



Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü

Bilgi ve Belge Yönetimi Anabilim Dalı

**ATIF KLASİKLERİNİN ETKİSİNİN VE İLGİLİLİK  
SIRALAMALARININ PENNANT DİYAGRAMLARI İLE ANALİZİ**

Müge Akbulut

Yüksek Lisans Tezi

Ankara, 2016

ATIF KLASİKLERİNİN ETKİSİNİN VE İLGİLİLİK  
SIRALAMALARININ PENNANT DİYAGRAMLARI İLE ANALİZİ

Müge Akbulut

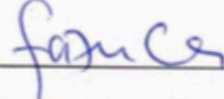
Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü  
Bilgi ve Belge Yönetimi Anabilim Dalı

Yüksek Lisans Tezi

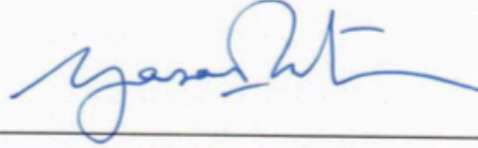
Ankara, 2016

## KABUL VE ONAY

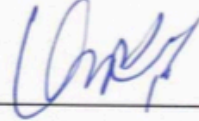
Müge Akbulut tarafından hazırlanan "Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi" başlıklı bu çalışma, 16 Haziran 2016 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından yüksek lisans tezi olarak kabul edilmiştir.



Prof. Dr. Fazlı Can (Başkan)



Prof. Dr. Yaşar Tonta (Danışman)



Doç. Dr. Umut Al

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Sibel BOZBEYOĞLU


Enstitü Müdürü

## BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin basılı ve elektronik kopyalarının Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin tamamı her yerden erişime açılabilir.
- Tezim sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin ..... yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

16 Haziran 2016



---

Müge AKBULUT

## TEŐEKKÜR

Bu alıŐma s¼recinde desteęini g¼rd¼ę¼m pek ok kiŐiye teŐekk¼r borluyum. En baŐta bana her zaman yol g¼steren ve d¼Ő¼nmemi saęlayan deęerli danıŐmanım, sevgili hocam Prof. Dr. YaŐar Tonta'ya sonsuz emekleri, sabrı ve desteęi iin ok teŐekk¼r ederim.

J¼rimde yer alarak deęerli g¼r¼Őlerini ileten Prof. Dr. Fazlı Can ile her defasında sorularıma sabırla cevap veren Do. Dr. Umut Al'a emeklerinden dolayı teŐekk¼r ederim.

Dikkatli okuması ve deęerli fikirleri ile alıŐmama katkıda bulunan, kendisinden ok Őey ¼ęrendięim ve ¼rnek aldıęım sevgili arkadaŐım S¼meyye Aka'ya minnettarım.

alıŐmam sırasında takıldıęım yerlerde kendisine danıŐtıęım, sorularımı ¼zenle cevaplayan, kendi yazdıęı sosyal aę analizi alıŐmalarında g¼rselleŐtirme iin kullanılan *CiteSpace* programına *Pennant diyagramları* ¼zellięini ekleyen Prof. Dr. Chaomei Chen'e teŐekk¼r ederim.

Tez konumun netleŐmesi s¼recinde bana zaman ayıran Esra ¼zkan elik'e ve alıŐmam sırasında kullandıęım kaynakların temin edilmesinde yardımını g¼rd¼ę¼m Cihan Doęan'a teŐekk¼r ederim.

Engin Almanca bilgisinden yararlandıęım ve her zaman yanımda olan Meri Dirik'e ok teŐekk¼r ederim. İlgi ve desteklerinden dolayı Sefa Dhyi, Derya Alptekin ve Zehra TaŐkın'a da teŐekk¼r borluyum.

Her zaman ve her koŐulda desteklerini hissettięim sevgili anneme ve babama hep yanımda oldukları iin minnettarım.

## ÖZET

AKBULUT, Müge. *Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi*, Yüksek Lisans Tezi, Ankara, 2016.

Atıf dizinleri bilim insanlarının ve bilimsel çalışmaların literatüre olan katkılarının ölçümüne yönelik otorite kaynaklardır. Bilgi erişim literatüründeki birçok çalışma erişim algoritmalarının geliştirildiği araştırmalara dayanmaktadır. Bilgi erişim alanının disiplinlerarası yapısı dolayısı ile bu çalışmalara birçok farklı alandan atıflar yapılmaktadır. Bilgi erişim literatüründe “klasik” olarak nitelendirilen ve birçok alanı etkileyen çalışmaların özellikle geriye dönük olarak incelenmesi önemlidir. Fakat özellikle eski tarihli çalışmaların etkilerinin atıf dizinlerinde gözlenmesi kolay değildir. Geleneksel atıf analizi çalışmanın kendi dönemindeki ve daha sonraki dönemlerdeki çalışmalar üzerindeki etkilerini ortaya çıkarmak için yeterli değildir. Bu çalışmaların diğer disiplinlere etkileri ve alandaki yeni modellere katkılarının ortaya çıkarılması büyük resmi görebilmek açısından önemlidir.

Bu çalışma kapsamında bilgi erişim literatüründe atıf klasiği haline gelmiş olan Maron ve Kuhns'un 1960 yılında yayımladıkları “olasılıksal erişim” ile ilgili çalışmanın literatürdeki etkisi ilgililik kuramı (relevance theory), bilgi erişim ve bibliyometriye dayanarak geliştirilen pennant diyagramları aracılığıyla görselleştirilmiştir. Bu amaçla temel hipotezi “Geleneksel atıf analizi ile gözlenemeyen disiplinlerarası ilişkiler pennant diyagramları yöntemi ile ortaya çıkarılabilir” şeklinde belirlenmiştir. Hipotezi test etmek için Maron ve Kuhns'un çalışmasına 1960 ile 2015 yılları arasında atıf yapan çalışmalar (toplam 4176 tekil çalışma) kaynakça bilgileri ile birlikte indirilmiş ve MS Excel programında yazılan makrolar yardımıyla hesaplamalar yapılarak grafikler hazırlanmıştır. Bu çalışmalardan kolayda örneklem yöntemi ile seçilen 90 çalışma için etkileşimli ve statik pennant diyagramları oluşturulmuş ve bu diyagramlar ayrıntılı olarak incelenmiştir.

Bu çalışmanın bir diğer önemli çıktısı da ilgililik sıralamalarıdır. Atıf dizinlerinde halihazırda kullanılmakta olan *kaynakça benzerliğine dayalı* ilgililik sıralamasına alternatif olarak çekirdek makalenin kaynakçası dışında diğer araştırmacıların atıflarının da hesaplamaaya dahil edildiği *pennant diyagramı yöntemi* ile ilgililik sıralaması oluşturulmuş ve bu sıralamalar birbirleri ile karşılaştırılmıştır. Bulgular hipotezleri destekler niteliktedir. Pennant diyagramları çekirdek makalede geçen olasılıksal modelin doğrudan ya da dolaylı olarak hangi modelleri etkilediği ya da hangi modellerden etkilendiği hakkında bilgi vermektedir. Çekirdek makalenin bilgi erişim alanı ve diğer disiplinlere katkıları ve birbirinden kopuk gibi gözükken alanlar arasındaki ilişkiler ile yazar, çalışma ve dergiler arasında belirli olmayan ve geleneksel atıf analizi ile belirlenemeyen ilişkiler ortaya çıkarılmıştır. Pennant diyagramları yöntemi kullanılarak oluşturulan ilgililik sıralamaları kaynakça benzerliğine göre oluşturulan ilgililik sıralamasından daha başarılı bulunmuştur. Çalışma Türkiye literatüründe ilgililik sıralaması oluşturulması kapsamında pennant diyagramları yönteminin kullanıldığı ilk çalışmadır. Bunun dışında pennant diyagramlarının etkileşimli versiyonu da ilk kez bu çalışmada hazırlanmıştır.

Çalışma kapsamında oluşturulan grafiklerde ve ilgililik sıralamalarında kullanılan veriler (toplam atıf ve ortak atıf sayıları) atıf dizinlerinde mevcuttur. Dolayısıyla atıf dizinlerinde kullanıcılara bu çalışmadakine benzer alternatif ilgililik sıralamaları sunulabilir. Pennant diyagramları hem araştırmacıların literatürü izlemelerini kolaylaştırılabilir hem de bir çalışmanın belli bir alanda ya da farklı alanlardaki çalışmalarını nasıl etkilediği belirlenebilir.

### **Anahtar Sözcükler**

Bilgi Erişim, Bibliyometri, Pennant Diyagramları, İlgililik, İlgililik Sıralamaları, Etki, Tf\*idf Ağırlıklandırma

## ABSTRACT

AKBULUT, Müge. *The Analysis of the Impact of Citation Classics and Relevance Rankings Using Pennant Diagrams*, Master's thesis, Ankara, 2016.

Citation indexes are important authority resources for measuring the contribution of scientists and scientific publications to literature. Many studies in information retrieval are based on research aiming to develop retrieval algorithms. These studies tend to receive citations from different fields because of the interdisciplinary nature of information retrieval. Therefore, it is important to analyze the so-called "citation classics" retrospectively to find out their impact on other fields. Yet, it is not easy to do this using citation indexes, especially for relatively old papers, as traditional citation analysis tends not to reveal the full impact of a work on other studies at its time and periods that follow. In order to see the big picture it is important to study the contribution of these studies on other disciplines as well.

In this study the impact of Maron and Kuhns' citation classic on "probabilistic retrieval" published in 1960 has been visualized using pennant diagrams that were developed on the basis of relevance theory, information retrieval and bibliometrics. We hypothesized that "The interdisciplinary relations that are unobservable with traditional citation analysis can be revealed using the pennant diagrams method". In order to test the hypothesis works that cited Maron and Kuhns' study between the years of 1960 and 2015 have been downloaded with their references (a total of 4,176 unique works) and graphics have been prepared by the macros written in MS Excel. Of 4,176 works, 90 were selected using convenience sampling techniques to create static and interactive pennant diagrams for further analysis.

Another important output of this study is the relevance rankings. As an alternative to the *relevance rankings based on the similarity of references* already used in citation indexes, relevance rankings have been created using the *pennant diagrams* that took into account not only items that cited the core



(seed) paper but also citations to the items that cited the core paper. Relevance rankings based on the similarity of references and that of pennant diagrams have been compared. Findings support the hypothesis in that pennant diagrams provide information as to which papers that the core paper on probabilistic model influenced or got influenced from, directly or indirectly. Relevance ranking based on pennant diagrams revealed the impact of the core paper on information retrieval field as well as on other disciplines. Furthermore, it identified the relations between these somewhat disconnected fields, between authors, works, and journals that cannot be readily identified using traditional citation analysis. Relevance rankings using pennant diagrams seem to have been more successful than the relevance rankings based on references similarity. This study is the first such study in Turkey that uses pennant diagrams for relevance rankings.

The data used in graphs and relevance rankings are available through citation indexes (the frequencies of total citations and co-citations). Thus, alternative relevance rankings based on pennant diagrams can be offered to users. Pennant diagrams can help researchers track the relevant literature more easily as well as identify how a core work influences other works in a specific field or in other fields.

### **Keywords**

Information Retrieval, Bibliometrics, Pennant Diagrams, Relevance, Relevance Rankings, Impact, Tf\*idf Weighting

## İÇİNDEKİLER

KABUL VE ONAY.....	ii
BİLDİRİM.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	v
ABSTRACT.....	vii
İÇİNDEKİLER.....	ix
ŞEKİLLER DİZİNİ.....	xi
TABLOLAR DİZİNİ.....	xii
KISALTMALAR DİZİNİ.....	xiii
1. BÖLÜM: GİRİŞ.....	1
1.1. KONUNUN ÖNEMİ.....	1
1.2. ARAŞTIRMANIN PROBLEMİ VE AMACI.....	1
1.3. ARAŞTIRMA SORULARI VE HİPOTEZLER.....	3
1.4. YÖNTEM.....	4
1.5. ARAŞTIRMANIN DÜZENİ.....	5
2. BÖLÜM: KAVRAMSAL ARKAPLAN VE LİTERATÜR DEĞERLENDİRMESİ.....	7
2.1. KAVRAMSAL ARKAPLAN.....	7
2.1.1. Giriş.....	7
2.1.2. İlgililik Teorisi.....	7
2.1.3. Pennant Diyagramları.....	12
2.1.4. İlgililik Teorisi, $Tf*Idf$ Ağırlıklandırması ve Pennant Diyagramları.....	16
2.1.5. Atıf Dizinlerinde İlgililik Sıralaması.....	18
2.2. LİTERATÜR DEĞERLENDİRMESİ.....	20
3. BÖLÜM: YÖNTEM.....	24
3.1. GİRİŞ.....	24
3.2. MARON VE KUHN'SUN (1960) ÇEKİRDEK MAKALESİ.....	24
3.3. VERİLERİN TOPLANMASI VE ANALİZE UYGUN HALE GETİRİLMESİ.....	25

3.3.1. Çekirdek Makale ile Birlikte Geçiş Sıklığının ( <i>tf</i> ) Hesaplanması.....	26
3.3.2. Toplam Atıfların Yıllara Göre Dağılımının Hesaplanması ...	28
3.3.3. Standardizasyon .....	29
3.3.4. Ortak Atıf ve Toplam Atıf Sayılarının Birleştirilmesi ve Etkileşimli Zaman Çizelgesinin Hazırlanması .....	30
3.3.5. Pennant Diyagramlarının Oluşturulması .....	32
3.3.6. Etkileşimli Zaman Çizelgesinin Oluşturulması .....	33
3.3.7. Kaynakça Benzerliğine Dayalı İlgililik Sıralamasının Oluşturulması.....	34
4. BÖLÜM: BULGULAR VE DEĞERLENDİRME .....	35
4.1. GİRİŞ .....	35
4.2. MARON VE KUHNS'UN (1960) ÇEKİRDEK MAKALESİ İÇİN PENNANT DİYAGRAMLARI .....	35
4.2.1. Pennant Diyagramlarının İlgililik Açısından Yorumlanması.....	38
4.2.2. Maron ve Kuhns'un (1960) Literatüre Etkilerinin Pennant Diyagramları ile İzlenmesi .....	44
4.2.3. Pennant Diyagramları ve İlgililik Puanları .....	51
4.3. MARON VE KUHNS'UN (1960) ÇEKİRDEK MAKALESİ İÇİN ETKİLEŞİMLİ PENNANT DİYAGRAMLARI.....	54
4.4. İLGİLİLİK SIRALAMALARI.....	55
5. BÖLÜM: SONUÇ VE ÖNERİLER .....	60
5.1. SONUÇ .....	60
5.2. ÖNERİLER.....	61
5.3. ÇALIŞMANIN SINIRLILIKLARI VE GELECEKTE YAPILMASI ÖNERİLEN ÇALIŞMALAR.....	62
KAYNAKÇA.....	64
EKLER.....	80

## ŞEKİLLER DİZİNİ

Şekil 1. Pennant diyagramı örneği.....	13
Şekil 2. WoS standart veri dosyası örnek gösterimi .....	28
Şekil 3. Ortak atıf saydırma makrosu sonucu .....	31
Şekil 4. Atıf raporu .....	31
Şekil 5. Maron ve Kuhns (1960) pennant diyagramı .....	36
Şekil 6. Maron ve Kuhns (1960) pennant diyagramı yoğunluk gösterimi .....	37
Şekil 7. Maron ve Kuhns (1960) pennant diyagramı (yazara göre) .....	39
Şekil 8. Başlığında “probabilistic” geçen çalışmalar.....	40
Şekil 9. Çalışmaların konusal olarak dağılımı .....	43
Şekil 10. Sorgu terimi <i>social</i> ile Venn diyagramı.....	46
Şekil 11. İlgililik sıralamalarının pennant diyagramı ile gösterimi .....	53
Şekil 12. Maron ve Kuhns (1960) etkileşimli pennant görüntüsü.....	54

## TABLolar DİZİNİ

Tablo 1. White, 2015 çekirdek makalesi için örnek veri seti .....	19
Tablo 2. İlgililik sıralamaları .....	20
Tablo 3. Sıklık verilerinin yıllara göre dağılımı .....	28
Tablo 4. Pennant diyagramındaki çalışmaların dağılımı .....	38
Tablo 5. WoS ilgili kayıtlar özelliğine göre oluşturulan listedeki kaynakların araştırma alanlarına göre dağılımı .....	57

## KISALTMALAR DİZİNİ

A&HCI	Arts and Humanities Citation Index
BKCI	Book Citation Index
CPCI	Conference Proceedings Citation Index
DF	Belge sıklığı (Document Frequency)
IDF	Ters belge sıklığı (Inverse Document Frequency)
ISI	Institute for Scientific Information
PY	Yayın yılı (Publication Year)
PFNETs	Pathfinder Ağları
RPYS	Referans Yılı Spektroskopisi (Reference Publication Year Spectroscopy)
SCI	Science Citation Index
SSCI	Social Sciences Citation Index
SOMs	Özörgütlemeli Haritalar (Self-Organizing Maps)
TF	Terim Sıklığı (Term Frequency)
WoS	Web of Science

# 1. BÖLÜM: GİRİŞ

## 1.1. KONUNUN ÖNEMİ

Herhangi bir bilimsel alanda dönemin belli başlı yazarlarının belirlenmesi, kendinden önceki çalışmalara referans vermemiş olsa bile yazarların hangi kaynaklardan yararlandıklarının ortaya çıkarılması o alanın araştırma örüntüsünün (pattern) belirlenebilmesi açısından önemlidir (Chen, 2006). Bir alanın entellektüel yapısını belirlemeye yönelik bu tür incelemeler genellikle atıf dizinlerinden (citation indexes) elde edilen ham atıf sayıları kullanılarak yapılmaktadır (Al, Sezen ve Soydal, 2012). Fakat bilim insanlarının ve bilimsel çalışmaların literatüre olan katkılarının ölçümünde ham atıf sayılarının kullanılması bu alandaki örüntüleri (patterns) tam olarak yansıtmamaktadır (Bauer ve Bakkalbasi, 2005; Ding ve diğerleri, 2014; Funkhouser, 1996; Jeong, Song ve Ding, 2014; Zhang, Ding ve Milojević, 2013).

Günümüzde atıf dizinlerinin performans değerlendirmede yoğun olarak kullanılması nedeniyle bilimsel çalışmalarda atıf yapılan kaynak sayısı da artmıştır (Taşkın, 2012). Fakat özellikle literatürdeki çok sayıda çalışmayı etkileyen eski tarihli "klasik"ler atıf dizinlerinde yeterince temsil edilmemektedir.

Atıf dizinleri atıf analizi aracılığıyla çalışmaların ve yazarların literatüre katkısının ölçülmesi dışında araştırmacıların literatür takiplerini kolaylaştırmak için de kullanılmaktadır. Bu bağlamda atıf dizinlerinde en çok kullanılan özellik ilgililik sıralamalarıdır. İlgililik sıralamaları yazarların arama yaptığı kaynağa benzer çalışmaları ya da benzer konuda çalışan diğer araştırmacıları saptamaları açısından önemlidir (Clark, 2013).

## 1.2. ARAŞTIRMANIN PROBLEMİ VE AMACI

Atıflar araştırmacıların çalışmalarını yaparken yararlandıkları kaynakları göstermektedir. Belli kaynakları referans listesinde bulunduran bir yazar atıfların niteliği negatif de olsa bu çalışmaları kendi çalışması ile ilgili bulmuştur. Dolayısıyla araştırmalar arasındaki ilişkiler belirlenirken en çok kullanılan

gösterge atıf sayısıdır (Cole, 2000). Fakat hem makale başına ortalama atıf sayısı hem de bir makalenin atıf alması için gereken zaman disiplinler arasında farklılık göstermektedir (Uçak ve Al, 2009). Bu, farklı alanlardaki makalelerin aldıkları atıf sayılarında çok büyük farklılıklara neden olmakta ve bu göstergenin disiplinlerin değerlendirilmesinde kullanılmasını zorlaştırmaktadır (Papp ve diğerleri, 2013). Bu durum disiplinlerarası bir çalışma alanı olan “bilgi erişim” (information retrieval) için de geçerlidir. Bu alandaki çalışmalara birçok farklı alandan atıflar yapılmaktadır. Fakat bilgi erişim alanındaki araştırma örüntülerinin belirlenmesi ve disiplinlerarası ilişkilerin gözlenmesi açısından geleneksel atıf analizi yetersiz kalmaktadır (Bauer ve Bakkalbasi, 2005).

Bu çalışmada *olasılıksal modelin* ortaya atıldığı ve bilgi erişim alanında “*atıf klasiği*”<sup>1</sup> sayılan bir çalışmanın (Maron ve Kuhns, 1960) literatüre olan katkısı *ilgililik teorisine dayanan pennant diyagramları* yöntemi kullanılarak yorumlanmıştır. Pennant diyagramlarında çalışmaların *bilişsel etkisi* (cognitive impact) *erişim kolaylığı* (ease of access) ile ilişkilendirilmekte ve daha iyi gözlenebilmektedir. Böylece Maron ve Kuhns’un (1960) bilgi erişim literatüründe ve diğer disiplinlerde hangi çalışmalarını etkilediği belirlenmiş ve bu çalışmalar ile yazarları arasındaki açık olmayan ilişkilerin ortaya çıkarılması sağlanmıştır. Bilgi erişim klasikleri dönemi öncesi ve sonrasında spesifik olarak farklı alanlarda çalışan araştırmacılar bir araya getirilmiştir. Hazırlanan etkileşimli zaman çizelgesi ile bu ilişkilerin zamana göre değişimi izlenebilmektedir. Araştırma kapsamında bilgi erişim tarihi açısından da bazı önemli bulgular elde edilmiştir.

Atıf dizinlerinin önemli özelliklerinden biri de ilgililik sıralamalarıdır (relevance rankings). Arama yapılan çalışma ile benzer nitelikteki kaynaklar çeşitli algoritmalar kullanılarak sıralanır. Atıf dizinlerinde ilgililik sıralaması çoğunlukla kaynakça benzerliğine göre yapılmaktadır. İki çalışmanın kaynakçası ne kadar fazla ortak kaynak içeriyorsa bu iki çalışmanın aynı konuda olma olasılıkları o kadar yüksektir. Ancak bu yöntem yeni bir yaklaşımın ortaya atıldığı

---

<sup>1</sup> Bir çalışmanın atıf klasiği sayılabilmesi için en az 100 atıf alması gerekmektedir (Fenton, Roy, Hughes ve Jones, 2002, s. 494). Çalışmada kullanılan Maron ve Kuhns (1960) ise Web of Science üzerinde 294 atıf almıştır.



çalıřmalarda ve kaynakçasındaki alıřma sayısı az olan ya da sz konusu alıřmaların dađılımları orijinal alıřma ile benzer olmayan kaynaklar iin bařarılı sonular getirmemektedir.

Pennant diyagramları ynteminde ise ham atıf deđerlerinin yanında ortak atıf deđerleri de hesaplamaya dahil edilerek ilgililik sıralamaları oluřturulmaktadır. Bu alıřmada ilgililiđi ortaya ıkarmak amacıyla alıřmalar arasındaki etkileřimleri her yazarın ekirdek<sup>2</sup> yazarlar (seed) ile ortak atıfları ve toplam atıfları kullanılarak pennant diyagramları aracılıđıyla grselleřtirilmiř ve yeni ilgililik sıralamaları oluřturulmuřtur.

Trkiye’de ilgililik sıralamaları ile ilgili az sayıda arařtırma bulunmaktadır (Ateř, zkan, Soysal ve Alatan, 2011; apkın, 2011; Emirzade ve Bitirim, 2008; Yıldırım ve Yeřilyurt, 2014). Bu alıřma ilgililik teorisine dayanan pennant diyagramları ynteminin uygulandıđı ilk alıřmadır. Bu ynyle yerel literatrde bařlangı niteliđindedir.

### 1.3. ARAřTIRMA SORULARI VE HİPOTEZLER

Bu arařtırmada yanıt aranan sorular řunlardır:

1. Maron ve Kuhns’un (1960) olasılıksal eriřim modeli bilgi eriřim literatrndeki hangi modelleri etkilemiř ya da hangi modellerden etkilenmiřtir?
2. Maron ve Kuhns’tan (1960) nce veya sonra olasılıksal modeller konusunda ve yakın konularda alıřanlar kimlerdir?
3. Etkileřimli pennant diyagramlarında olasılıksal eriřim modelinin hangi yıllar aralıđında hangi modelleri etkilediđi gzlenebilir mi?
4. Pennant diyagramları yntemi ile elde edilen *ilgililik sıralaması* ile Web of Science veri tabanındaki ilgili kayıtlar (related records) zelliđi ile elde edilen ilgililik sıralaması ne kadar rtřmektedir?

---

<sup>2</sup> ekirdek terimi, pennant diyagramları ynteminde literatrdeki etkisi belirlenmek istenen ya da ilgililik sıralaması oluřturulan alıřma (veya yazar) anlamında kullanılmaktadır.

Bu araştırmanın temel hipotezi “Geleneksel atıf analizi ile gözlenemeyen disiplinlerarası ilişkiler pennant diyagramları yöntemi ile ortaya çıkarılabilir” şeklinde belirlenmiştir. Bu ilişkilerin saptanması alanın entellektüel yapısının doğru olarak belirlenmesi açısından önemlidir. Bir diğer hipotez ise “Ortak atıf sayılarına göre oluşturulan ilgililik sıralamaları kaynakça benzerliğine göre oluşturulan ilgililik sıralamasından daha başarılıdır” şeklinde oluşturulmuştur. Araştırmacılar arama yaptıkları konu ile ilgili konularda çalışan diğer araştırmacıları ya da benzer çalışmalarını saptarken ilgililik sıralamalarından yararlanmaktadır. Dolayısıyla ilgililik sıralamaları araştırmacıların literatür izlemelerini kolaylaştıran en önemli unsurdur.

#### 1.4. YÖNTEM

Pennant diyagramlarının ve ilgililik sıralamalarının oluşturulmasında kullanılacak veriler için Bilimsel Bilgi Enstitüsü (*ISI*) bünyesindeki Web of Science (*WoS*) veri tabanında *SCI* (fen bilimleri), *SSCI* (sosyal bilimler), *AHCI* (sanat ve insani bilimler), *CPCI* (konferans bildirileri) ve *BKCI* (kitaplar) atıf dizinleri taranmıştır. Pennant diyagramlarının hazırlanması için çekirdek makaleye atıf yapan çalışmalar kaynakça bilgileri ile birlikte indirilmiştir (herhangi bir yayın türü kısıtlaması yapılmamıştır). Yıl aralığı çekirdek makalenin yayınlandığı yıl olan 1960 ile 2015 arası olarak belirlenmiştir. 10 Aralık 2015 tarihinde yapılan taramaya göre Maron ve Kuhns’a (1960) 294 kere atıf yapılmıştır. Söz konusu çalışmalara atıf yapan çalışmaların referanslarının sayısı ise 9607’dir (tekil sayı ise 4176’dır). Bu sayıya kaynakçasında Maron ve Kuhns (1960) bulunmasa bile başlık, özet ve anahtar kelimelerinde *olasılıksal model* ya da *model 1* terimleri geçen kaynaklar da dahildir. Çalışma kapsamında 4176 çalışmadan ayrıntılı olarak incelenmek üzere *kolayda (kolaycı) örneklem* seçme tekniği<sup>3</sup> ile 90 çalışma (çekirdek makale hariç) seçilmiştir. Diyagram, işleme kolaylığı ve bilişsel etki değerleri dikkate alınarak 10 alana ayrılmıştır. Bu alanlarda yer alan çalışmaların birbirlerine yakınlıkları, çalışmaların konuları ve yazarlarının yayın

<sup>3</sup> Kolayda örneklem seçme yönteminde araştırmacı veri toplamak amacıyla kolayca erişebildiği deneklere başvurmaktadır.

yaptıkları alanlar içerik analizi yöntemi kullanılarak araştırılmıştır. Ardından 4176 çalışma için pennant diyagramları yöntemi ile ilgililik sıralaması oluşturulmuştur.

İlgili kaynağa atıf yapan kaynak WoS'ta dizinlenmemişse dış kaynak olarak adlandırılmaktadır. WoS ilgili kayıtlar sıralamasında dış kaynaklar yer almamaktadır. WoS ilgili kayıtlar sıralaması ve pennant diyagramları yöntemi ile oluşturulacak ilgililik sıralamaları arasında karşılaştırma yapılabilmesi için dış kaynak verileri bu çalışmaya dahil edilmemiştir. Dolayısıyla dış kaynaklar pennant diyagramlarının kapsamı dışındadır.

Yöntem ile ilgili daha ayrıntılı bilgi Bölüm 3'te verilmiştir.

## 1.5. ARAŞTIRMANIN DÜZENİ

Araştırma beş bölümden oluşmaktadır:

Birinci bölümde çalışmaların literatüre etkisinin izlenmesinin ve ilgililik sıralamalarının önemine ve bu işlemlerin atıf dizinleri üzerinden niçin sağlıklı bir şekilde yapılamadığına değinilmiştir. Bunun yanı sıra araştırmanın amacı, sınırlılıkları, araştırma soruları ve hipotezlere yer verilmiştir.

İkinci bölümde ise çalışmanın temelini oluşturan pennant diyagramları ile ilgililik teorisi, *tf\*idf* ağırlıklandırma hakkında bilgi verilmiş ve literatürdeki çalışmalar değerlendirilmiştir. Bunun dışında pennant diyagramlarının yorumlanmasına ve atıf dizinlerinde ilgililik sıralamalarının oluşturulmasına değinilmiştir. Çalışma kapsamında çekirdek makale olarak incelenen bilgi erişim literatüründe atıf klasiği haline gelmiş olan Maron ve Kuhns'un (1960) olasılıksal erişim konulu çalışması ile ilgili genel bilgiler de bu bölümde yer almaktadır.

Üçüncü bölümde pennant diyagramlarının ve ilgililik sıralamalarının hazırlanması ile ilgili olarak verilerin toplanması, analize uygun hale getirilmesi ve grafiklerin oluşturulması sürecinde izlenen yöntemler ve teknikler hakkında bilgi verilmektedir.

Dördüncü bölümde araştırma sorularını cevaplamak ve hipotezleri test etmek amacıyla çekirdek makale için oluşturulan pennant diyagramları ve ilgililik

sıralamalarına ilişkin bulgular sunulmuştur. Ardından *kaynakça benzerliğine dayalı* ilgililik sıralaması ile çekirdek makalenin kaynakçası dışında diğer araştırmacıların atıflarının da hesaplamaaya dahil edildiği *pennant diyagramı yöntemi* ile oluşturulan ilgililik sıralaması karşılaştırılarak bu iki sıralamanın birbirine benzeyip benzemediği tartışılmıştır.

Beşinci bölümde ise çalışma kapsamında elde edilen bulgular değerlendirilmiştir. Atıf dizinlerinde mevcut olan veriler kullanılarak arama yapılan bir çalışma ile ilgili daha isabetli ilgililik sıralamaları oluşturulması ve çalışmaların diğer çalışmalar ile ilişkilerinin gözlenmesi için diyagramlar oluşturulması için önerilerde bulunulmuş ve gelecekte yapılabilecek çalışmalar belirtilmiştir.

## 2. BÖLÜM: KAVRAMSAL ARKAPLAN VE LİTERATÜR DEĞERLENDİRMESİ

### 2.1. KAVRAMSAL ARKAPLAN

#### 2.1.1. Giriş

Bilgi erişim kullanıcının bilgi ihtiyacını tanımladığı sorgudaki terimler ile belgelerde geçen terimlerin eşleştirilmesine dayanmaktadır. Terimlerin eşleşme oranına göre sorgu-belge ve belge-belge benzerlikleri hesaplanır. Fakat eşleşme süreci tam olarak anlaşılmadığından “bilgi erişim sorunu”na henüz kesin bir çözüm bulunamamıştır (Tonta, 2001). Çünkü bilgi erişimde bazı belirsizlikler söz konusudur. Bunlardan en önemlileri belge ve sorgu temsili belirsizliğidir. Temsil için belirlenen terimler öznel; dolayısıyla kişiye, zamana ve duruma göre değişebilir. Bunların dışında bilgi erişim fonksiyonu konusunda da belirsizlik bulunmaktadır (Turtle ve Croft, 1997). Buradaki belirsizlik aynı kavramın farklı biçimlerde temsil edilebilme olasılığından kaynaklanmaktadır. Bu belirsizliklerin çözümüne yönelik bilgi erişim kuralları ve modeller geliştirilmiştir. Söz konusu kural ve modeller bilgi erişim sistemlerinde sorgu-belge ve belge-belge benzerliklerinin hesaplanmasına ve ilgililik sıralamaları oluşturulmasına olanak sağlamaktadır.

İdeal bilgi erişim sistemi ilgili belgelerin tümüne ve salt ilgili belgelere erişim sağlamalıdır. Yukarıda bahsedilen belirsizliklerden dolayı ideal bir bilgi erişim sistemi yoktur, fakat kullanıcılar genellikle birkaç ilgili belgeye fazla çaba harcamadan eriştikleri zaman tatmin olmaktadır (Tonta, 1995).

#### 2.1.2. İlgililik Teorisi

İlgililik (relevance) genelde bilgibilim, daha spesifik olarak bilgi erişim açısından oldukça önemlidir (Wilson, 1973). Öznel olduğu için ölçülmesi oldukça zor olan bu kavramı bilgibilimciler yarım yüzyıldan fazla süredir tartışmışlardır (Saracevic, 1975, Tonta, 2012). *İlgililik teorisi* temel olarak dinleyicilerin (hearers) konuşmacıların (speakers) söylediklerinden ne anladıklarını,

okuyucuların yazarların yazdıklarının altında yatan anlamları nasıl anladıklarını açıklamak amacıyla kullanılmıştır (Clark, 2013, s. 114).

Sperber ve Wilson (1995, s. 261) ilgililiği “insan aklının girdileri” olarak tanımlamışlardır. Girdiler hem kişinin sahip olduğu bilinci ve muhakeme yoluyla edindiği bilgileri hem de doğadan ve diğer kişilerden gelen iletileri içermektedir. Sperber ve Wilson bir girdinin ilgililiğini belirleyen şeyin o girdinin “bilişsel etki”si (cognitive effect) ile o girdiyi işlemek için gereken çaba (ease of processing) olduğunu vurgulamaktadırlar (Wilson ve Sperber, 2002)

İlgililik teorisinde yeni girdiler çeşitli bilişsel değişimlere yol açmaktadır. Girdiler mevcut bilgiyi (information) ve bilgi yapılarını (knowledge structures) değiştirebilecek ya da yeni bir sonuca ulaştıracak türde bir etkiye sahip olabilir. Girdiler mevcut bir varsayımı güçlendirebilir ya da eleyebilir. Dolayısıyla, kişinin var olan varsayımlarını değiştiren birer çıkarım olarak düşünülebilir (Clark, 2013; White, 2015). Girdiler varsayım ile birleşerek yeni bir sonuca ulaşmaya yol açarsa bu “pozitif<sup>4</sup> bilişsel etki” anlamına gelmektedir (Buchanan ve O’Connell, 2006; Sperber ve Wilson, 1995, s. 265-266).

Sperber ve Wilson’ın ilgililik teorisinde ilgililik bir oran olarak tanımlandığı için bilişsel etkilerin ve işleme çabasının ordinal olarak (sıralama ölçeği ile) ölçülebilmesi mümkündür. Fakat bu oran öznel ve kişilere, zamana ve duruma göre değişmektedir. İlgililik oranının hesaplanması için kullanılan formül aşağıda verilmektedir.

$$\text{“İlgililik} = \text{bilişsel etki} / \text{erişim kolaylığı (işleme çabası)} \text{”} \quad (1)$$

Formül 1’e göre bir girdinin ilgililiği bilişsel etkileri ile doğru orantılı, bu girdiyi işlemek için gerekli çaba ile ise ters orantılıdır. Yani işleme çabasındaki daha fazla maliyet ilgililiği düşürmektedir. İlgililik teorisinde anlamlar yorumlanırken kullanıcıların ilgililik beklentileri, ön yargıları ve varsayımları ön plandadır.

<sup>4</sup> “Pozitif” terimi memnun edici anlamına değil; doğru, gerçeğe dayalı anlamına kullanılmıştır.

Bilgi erişimde kullanıcılar erişim çıktılarının sorgu ile ilgililiğine karar verirken en erişilebilir (en az çaba gerektiren) yorumu ararlar (Soll, Milkman ve Payne, 2015). Genelde erişim kolaylığı belgedeki terimler (yani anlam) ile sorgudaki terimlerin ne derece eşleştiği ile ilgilidir (White, 2007a). Dolayısıyla terim eşleşmesine dayanan çıkarımlar kullanıcılar için en kolay olanıdır. Bu sebeple de sorgu ve dizin terimleri arasındaki eşleşme ya da benzerlik bilgibilimde konusal ilgililiğin (topical relevance) de temelini oluşturur (Carevic ve Schaer, 2014; White, 2007b). Terimler ile eşleşmeyen fakat ilgili olan belgeleri işlemek ise daha zordur. Çünkü sorgu terimleri ve erişim sonuçlarının eşleşme oranı çok yüksek olmasa bile konusal olarak ilgililik olabilir. Bunları işleyebilecek altyapıya sahip kişiler genelde uzmanlaşmış araştırmacılarıdır.

İlgililik ile ilgili bir diğer önemli kavram da çıkarımdır. Çıkarımsal (inferential) model, dinleyiciler tarafından anlaşılanların konuşma ya da yazının gerçekte içerdiğinden daha fazlası olduğu ile ilgilidir (Saracevic, 1996, s. 204). Konuşmacı ve yazarlar dinleyicilerin çıkarımsal yeteneklerine ve kavramsal arka plan bilgilerine güvendikleri için her şeyi açıklama gereği duymazlar (Clark, 2013, s. 86). Çıkarımsal modelde iletişim bu şekilde gerçekleşir. White (2007a, 2007b), pennant diyagramları yaklaşımında ortak atıf sayıları ve toplam atıf sayılarına bakarak ilgililik ile ilgili çıkarımlarda bulunmaktadır.

Bilgi erişim kuralları ve modellerinin temelinde de ilgililik teorisi bulunmaktadır. Erişim fonksiyonu modellerinden kesin çakışma (exact match) kuramına dayanan *Boole modelinde* sorgu çalıştırıldığında aranan terim ve belge arasında çakışma varsa belge "ilgili", yoksa "ilgisiz" olarak değerlendirilmektedir (ikili [binary] ağırlıklandırma).

Tam çakışmaya dayalı Boole modelindeki ikili ağırlıklandırma yerine *vektör uzayı modelinde* kısmi çakışmaya dayanan benzerlik oranı hesaplanmaktadır. Vektör uzayı modelindeki kısmi çakışma hem sorgu terimlerinin hem de belgelerdeki dizin terimlerinin ağırlıklandırılması ile gerçekleştirilmektedir. *Vektör uzayı modelinde* ağırlıklar hesaplanırken bir terimin belgede geçme sıklığı (term frequency) ve derlemdeki (koleksiyon) belgelerde geçme sıklığı (document frequency) kullanılmaktadır. Bir terimin belgedeki geçiş sıklığı

belgenin hem belirli bir konu ile olan ilgililiğini hem de derlemdeki diğer belgelerle olan yakınlığını belirlemektedir (Sparck Jones, 1972, 1973). Belgede geçen her terim bir vektör olarak çoklu uzayda temsil edilmektedir. Sorgu vektörü ile belge vektörü arasındaki açının büyüklüğüne bakılarak sorgunun hangi belgeyle daha ilgili olduğu belirlenmektedir.

*Tf\*idf* (*term frequency \* inverse document frequency*) değerleri bir belge uzayı içinde verilen bir kelimenin herhangi bir belgede ne kadar önemli olduğunu belirten istatistiksel bir ölçüdür. Terim sıklığı (*tf*) değeri terimin ilgili belgede kaç kez geçtiğini göstermektedir. Belge sıklığı (*df*) ise terimin derlemde geçtiği belge sayısıdır. *tf* değeri yüksek olan bir terim diğer belgelerde de sık geçiyorsa ayırt edici özelliği düşüktür. Öte yandan, diğer belgelerde seyrek geçen ve *tf* değeri yüksek olan terimin ayırt edici özelliği ise yüksektir (Glushko, 2015, s. 714; Manning ve Niyak, 2015; Stanford, 2012; Tonta, Bitirim ve Sever, 2002). Sparck Jones (1973) derlemdeki belgelerde sık geçen terimleri ağırlıklandırmak için *ters belge sıklığı* (*inverse document frequency*) değerinin kullanılmasını önermiştir. Dolayısıyla bir belgede sık geçen fakat derlemdeki belgelerde de sık geçen terimlerin ağırlıkları azaltılmıştır.

Bir terimin ilgili belgedeki ağırlığını hesaplamak için *tf\*idf* formülü kullanılmaktadır. Formüldeki ters belge sıklığı (*idf*) hesaplaması için de *log (N/df)* formülü kullanılmaktadır. Buradaki *N*, derlemdeki toplam belge sayısı, *df* ise ilgili terimin belge sıklığıdır. Ters belge sıklığı yaklaşımı derlemde daha az geçen ve ilgili bir belgede sık geçen terimleri göreceli ağırlığı daha yüksek olarak değerlendirir (Manning ve Niyak, 2015; Shah, 2009). Günümüzde ters belge sıklığı halen birçok arama motoru tarafından sorgu terimlerini ağırlıklandırma ve belgeleri sıralama amacıyla kullanılmaktadır.

*Tf\*idf* değerleri belge erişim sistemlerinde sorgu terimleri ve dizin (*index*) terimlerinin ağırlıklandırılması amacıyla da kullanılmaktadır (Manning, Raghavan ve Schutze, 2008). Belge erişim sistemlerinde ağırlıklandırılan kelimeler genellikle konusal olduğu için buradaki ilgililik, belgeler ile kullanıcıların sorgularının konusal olarak ne kadar eşleştiği ile ilgilidir.



Literatürde *Model 1* olarak da adlandırılan *olasılıksal modelde*; belgelerin iki değerli (ilgili ya da ilgisiz) olarak dizinlenmesi yerine; terimlerin ilgili belgelerde bulunabilme olasılıkları temel alınarak ağırlıklandırma yapılmaktadır. Belli bir belgenin bir sorgu ile ilgili olma olasılığının tahmin edilmesine dayanan modelde erişilen belgeler ilgililik olasılığına göre sıralanır (*olasılık sıralama ilkesi: probability ranking principle*) (Robertson, 1977).

Olasılıksal modelde belgeleri dizinleyen kişi geleneksel modelde olduğu gibi bir belgeye atanacak konu başlıklarına, anahtar kelimelere vs. karar verirken belge belli bir konu ya da anahtar kelimeyle “ilgili” veya “ilgili değil” şeklinde ikili (binary) karar vermez. Bunun yerine dizinci, bilgi ihtiyacını belli bir konu ya da anahtar kelimeyle ifade eden kullanıcının belli bir belge ile karşılaştığında bu belgeyi ilgili bulma olasılığına göre belgeye dizin terimi atar (Bookstein, 1983; Maron, Kuhns ve Ray, 1958, s. 45). Bir belgenin belli bir anahtar kelimeyle ifade edilen bilgi ihtiyacını karşılama olasılığı 0 (ilgisiz) ile 1 (ilgili) arasında değişmektedir. Örneğin, bir belge belli bir konudaki bilgi ihtiyacını daha çok (diyelim ki %80 -veya 0,8 oranında-, bir başka konudaki bilgi ihtiyacını ise daha az (%50 -veya 0,5) oranında karşılıyor olabilir. Başka bir deyişle, söz konusu belgenin ilk konu için ilgililik düzeyi 0,8, ikincisi için 0,5'tir.

Bir bilimsel alana en çok katkı sağlayan yazarların ve çalışmaların saptanmasını amaçlayan etki belirleme çalışmaları konusunda geniş bir literatür mevcuttur. Etki belirleme amacıyla en çok kullanılan yöntem *atıf analizi* ve *haritalamadır* (Sylvia, 1998). Özellikle atıf dizinleri yaygınlaşıp veri kaynağı olarak kullanılmaya başladıktan sonra atıf analizi çalışmalarında ciddi bir artış gözlenmiştir (Düzyol, 2011). Atıf analizi çalışmalarında en yaygın kullanılan teknikler herhangi iki çalışmaya başka çalışmalar tarafından birlikte atıf yapılma sıklığının hesaplandığı *ortak atıf analizi* ve iki farklı çalışmanın üçüncü bir çalışmaya ortak atıf yapma sıklığının hesaplandığı *bibliyografik eşleşmedir* (Al, 2008; Düzyol, 2011; Larsen, 2004, White ve McCain, 1998). Her iki yöntem çalışmaların birbirine benzerliği ve ilgililiği hakkında fikir verdiği için öneri sistemlerinde (recommendation systems) sıklıkla kullanılmaktadır (Beel, Gipp, Langer ve Breitinger, 2015; Carevic ve Mayr, 2014; McNee ve diğerleri, 2002;

Strohman, Croft ve Jensen, 2007). *Yazar ortak atıf analizinde* ise bir yazar çifti ne kadar fazla birlikte atıf alıyorsa, aralarındaki ilişkinin o derece yüksek olduğu düşüncesinden yola çıkılarak yazarların birbirlerine yakınlıkları hesaplanmaktadır (Düzyol, 2011; White, 2003; White ve Griffith, 1981, Zhang, Ding ve Milojević, 2013).

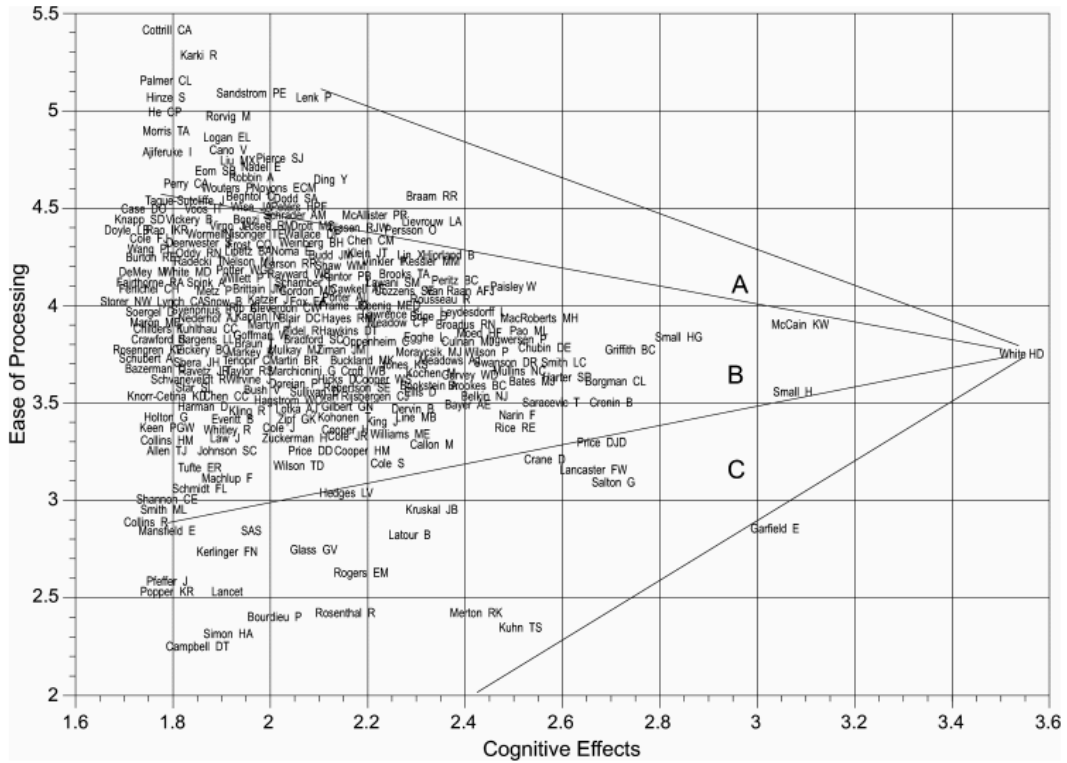
### 2.1.3. Pennant Diyagramları

İlk kez Howard White (2007a, 2007b) tarafından ortaya atılan *pennant diyagramları* yaklaşımının temeli *ilgililik teorisi (relevance theory)*, *tf\*idf* ve *bibliyometriye* dayanmaktadır. White çalışmalar ve yazarlar arasındaki ilgililiği, *erişim kolaylığı* ve *bilişsel etki* ile ilişkilendirerek *vektör uzayı modelindeki tf\*idf'i konusal ilgililik ve bilginin elde edilmesi (erişim kolaylığı)* olarak *x* ve *y* eksenlerine yansıtıp benzer sonuçlar elde etmiştir.

Pennant diyagramları terimleri ağırlıklandırmak için kullanılan *ters belge sıklığı (idf)* yöntemi ile yakından ilgilidir. Pennant diyagramları yönteminde *tf* ve *idf*, belgeleri sıralamada sorgu terimlerini ağırlıklandırma yerine bibliyometrik dağılımlardaki terimleri ağırlıklandırmada kullanılmaktadır. White'ın çalışmalarında (2007a, 2007b, 2009, 2010, 2015, 2016) tamamen kelime sıklığına dayanarak oluşturulan diyagramların konusal ilgililik ve o bilginin ele geçirilmesi için ne kadar çaba harcanmasıyla örtüştüğü görülmektedir.

Pennant diyagramının örnek bir gösterimi Şekil 1'deki gibidir. Her bir yazarın çekirdek yazar ile ortak atıf sayılarından elde edilen değerler *x* ekseninde yer almakta ve bilişsel etkiyi belirtmektedir. *Y* ekseninde yer alan *idf* değeri ise her bir çalışmanın tek başına aldığı atıf sayılarından elde edilmekte ve makaleye erişim kolaylığını vermektedir. Pennant diyagramlarında iki boyutlu saçılım grafiği çizilirken çekirdek terim ya da yazarlar *tf\*idf* ağırlığına göre sıralanmaz. Bunun yerine terim ya da yazarların normalleştirilmiş *tf* ve *idf* değerleri diyagrama ayrı ayrı yerleştirilir. Diyagramlar yorumlandığında yazarların uzmanlık düzeyleri, yazarlar ve çalışmalar arasındaki konusal ilgililik, metinlerin ve yazarların otoritesi, mesajların işleme çabası ve bilişsel etki hakkında bilgi sağlanmaktadır. Bunun dışında ilgililik teorisinde erişim kolaylığı oranı olarak

tanımlanan ilgililik ve terimlerin spesifikliđi (işlenme çabasını etkilemektedir) ile ilgili de bilgi sağlanmaktadır (White, 2007b, 2015).



Şekil 1. Pennant diyagramı örneđi (White, 2007a, s. 551)

### 2.1.3.1. $TF*IDF$ ve Pennant Diyagramları

$Tf*idf$  yaklaşımı bilgi erişim literatüründe daha çok doğal dildeki konusal isim öbekleri için kullanılmaktadır. White ise bu yöntemi ortak atıf alan yazarlar ve çalışmalara da uygulamıştır. Aslında, yazarlar için de konusal terimlerde olduğu gibi bir durum söz konusudur. Örneđin, pek çok matematikçi *Cahit Arf* adının *Arf halkaları*, *Arf deđişmezi* ve *Hasse-Arf teoremini* temsil ettiđini ya da ilgili teoremlerin *Cahit Arf*ı temsil ettiđini bilir. Bu bağlamda “Cahit Arf” konusal bir terim olarak düşünülebilir.

Yüksek *terim sıklığı* değeri ağırlıkları belli bir sorgu teriminin göreceli olarak daha sık görüldüğü belgeleri ilgililik açısından ön plana çıkartır. Kullanıcı ilgili terimi erişim sonuçlarında gördüğünde pozitif kavramsal etki oluşur ve sonuçların sorgu ile ilgili olduğu varsayımı güçlenir. Öte yandan, yüksek *ters belge sıklığı*, sorgu ile daha ilgili olan terimleri içeren belgeleri ve kullanıcının

işlemesi için daha kolay olduğu varsayılan belgeleri ağırlıklandırarak erişim listesinde daha yukarıda gösterir (White, 2016). Bu durum ilgililik oranı yüksek olup bilginin elde edilmesi için harcanan çabanın düşük olması anlamına gelmektedir.

### 2.1.3.2. Pennant Diyagramlarının Yorumlanması

Pennant diyagramları yönteminde belgelerde geçen terimlere ait *tf* ve *idf* değerleri *x* ve *y* eksenlerine ayrı ayrı yerleştirildikleri için belgeler sadece *tf\*idf* ağırlıklarına göre sıralandığında gözlenmesi mümkün olmayan ilişkiler saçılım grafiğinde kolaylıkla gözlenebilmektedir. Dolayısıyla pennant diyagramları disiplinlerin entellektüel yapılarının izlenmesi dışında yazarlar ve çalışmalar arasındaki konusal ilgililik, işleme çabası, bilişsel etki gibi konularda da öngöründe bulunmamızı sağlar. *Roget* ve *WordNet* gibi kavram dizinlerinde (thesaurus) görülen türde geleneksel semantik hiyerarşiler net olmasa da, pennant diyagramlarında oluşan geniş gruplar niteliksel sınıflamalara olanak sağlamaktadır (McHale, 1998; White, 2015).

*Bilişsel etki (x)* eksenindeki değerler ortak atıf verilerinden oluşmaktadır. Ortak atıfı ilgililik teorisi bağlamında yorumlayacak olursak: İki çalışma her birlikte anıldıklarında pozitif kavramsal etki üretirler ve bu iki çalışmanın ortak atıf değeri ne kadar yüksekse ikisi arasındaki bağlantıyı yakalamak o kadar kolaydır. Aynı şekilde iki yazarın ortak atıf sayısı yüksekse bu yazarların aynı alanda çalışma olasılıkları yüksektir. Diyagramdaki konumlar ortak atıf bağlamında incelendiğinde; bir çalışmanın toplam atıf sayısı ve çekirdek makale ile birlikte aldığı atıf sayısı birbirine yakınsa, Şekil 1'deki pennant diyagramındaki bir nokta *x* ekseninde çekirdek makaleye doğru yaklaşacaktır. İlgililik açısından *x* ekseni üzerinde çekirdek makaleye daha yakın olan çalışmalar çekirdek makale ile daha ilgilidir. En uzak noktadaki (en soldaki) çalışmalar ise en az ilgili olanlardır.

Pennant diyagramlarında *x* ekseninde en soldaki noktalar çekirdek çalışma ile sadece bir kez birlikte atıf almış çok sayıda çalışmadan oluşmaktadır. Sağa doğru ilerledikçe çalışmalar sayıca azalmaktadır. Bunun nedeni çekirdek

makale ile anılma sayısı ve çalışma sayısı arasında ters orantı oluşudur. Dolayısıyla bilişsel etki ekseninde sağa doğru gittikçe noktalar en ilgili değer (çekirdek makale) olana kadar sayıca azalmaktadır. Bir flamaya benzeyen bu diyagramlar “pennant diyagramı” olarak adlandırılmıştır. Bu özellik daha çok ya da daha az sayıda belgenin benzer *tf* ve *idf* değerlerine sahip olmasıyla ilgilidir (bkz. Şekil 1).

Y ekseninde gösterilen erişim (işleme) kolaylığı ise herhangi bir çalışma ile çekirdek makale arasındaki bağlantının ne kadar kolay görülebildiği ile ilgilidir. Diyagramın en tepesindeki noktalar çekirdek makale ile ilgililikleri açısından işlemesi kolay olanlardır, daha aşağıda olanlar ise çekirdek makaleyle göreceli olarak daha az ya da dolaylı olarak ilgilidir.

Bilişsel etki skalasının yüksekliği ya da düşüklüğü atıf yapanların yargıları ile belirlenir. Çünkü atıflar, araştırma sırasında hangi kaynaklardan yararlandığı hakkında bilgi vermektedir (Al ve Tonta, 2004). Dolayısıyla atıflar ilgililik belirleme açısından en önemli unsurlardan birisidir.

Pennant diyagramlarında toplam atıf (*df*) değerleri *y* eksenine yansıtılırken  $N/\log(idf)$  formülü kullanıldığı için payda büyüdükçe normalleştirilmiş *idf* değeri küçülmekte ve ilgili noktanın konumu aşağı doğru inmektedir. *Idf* değeri çok yüksek olan çalışmalar pennant diyagramında baskılanmakta ve bu çalışmaların ağırlıkları düşmektedir. Dolayısıyla *idf* değeri yüksek olan bir çalışmanın *y* ekseninde nispeten aşağı bölgelerde olması beklenmektedir.

Pennant diyagramı yorumlanırken öncelikle çekirdekte yer alan yazar ya da çalışmalardan yola çıkılarak, diyagramdaki noktalar üç gruba ayrılmaktadır (bkz. Şekil 1). Bu üç gruptan A bölgesinde yer alan yazarların çekirdekte yer alan yazarın ardılları, B kesiminde yer alan yazarlar çekirdekteki yazarın akranları, C kesiminde yer alan yazarların ise çekirdekte yer alan yazarın öncülleri olması beklenmektedir. A kesimindeki yazarların çekirdek yazarla ilgisinin net olması beklenmektedir. C kesimine doğru inildikçe bu bağlantıların kurulabilmesi için alandan uzman görüşüne ihtiyaç duyulmaktadır. Yazarların yaşlarının da bu ilişkiye göre yukarıdan aşağıya artması beklenmektedir. Ayrıca diyagram

üzerinde yer alan yazarların, soldan sağa doğru çekirdek yazara yaklaşırken bilişsel olarak çekirdek yazar ile daha fazla ilişkili olması beklenmektedir (Sperber ve Wilson, 1995; Tonta ve Özkan Çelik, 2013; White, 2007a, 2007b, 2009, 2010, 2011, 2015; White ve Mayr, 2013).

A sektöründe yer alan çalışmalar daha spesifik konularda yapılan çalışmalardır (genelde makaleler). Bu sektördeki çalışmalar çekirdek makalenin yazarının ya da ardıllarının konuyla ilgili yazdıklarından oluşmaktadır. B sektöründe daha çok çekirdek makale üzerine inşa edilen ama A sektöründekiler kadar spesifik ve çekirdek makalenin konusu ile doğrudan ilgili olmayan çalışmalar yer almaktadır. C sektöründeki çalışmalar ise daha genel konularda hazırlanmış rehber niteliğinde olan ve bilgilim literatüründe sıklıkla tercih edilen çalışmalardan oluşmaktadır. Bu çalışmalar genellikle daha eski tarihlidir.

Birikmiş üstünlük anlamına gelen *Matthew etkisi* (Matthew effect) sosyolojide zenginin daha zengin, fakirin daha da fakir olduğu durumları açıklamak için kullanılmaktadır. Matthew etkisi bibliyometride ise benzer nitelikte çalışmalar yapan iki bilimcinin, alanda tanınmış olanının daha az tanınana kıyasla çok daha yüksek atıf alacağı anlamına gelmektedir (Merton, 1968, 1988; Smucker, 2008). Yazarların atıf yapma eğilimleri baz alındığında genelde bir çalışmanın çok atıf alması bundan sonra da çok atıf alacağıyla ilgili bir göstergedir. Literatürde sıklıkla rastladığımız gibi, çalışma ile doğrudan ilgili olmasa da bazı çalışmalara tarihsel önemi nedeniyle yüksek sayıda atıf yapılmaktadır. Bu durum atıf tabanlı ölçümlerde sorunlar yaratabilmektedir (Wang, 2014). Bu tür çalışmalar pennant diyagramlarında C sektöründe yer almaktadır.

#### **2.1.4. İlgililik Teorisi, $TF*IDF$ Ağırlıklandırması ve Pennant Diyagramları**

İlgililik kuramının bilgilim açısından ve daha spesifik olarak bilgi erişim açısından önemli olduğu daha önce vurgulanmıştı. Harter (1992, s. 612–613) ilgililik teorisini belge erişiminde ilgililik yargılarının dışında bibliyometri ile de ilişkilendirmiştir. White ise ilgililik teorisini pennant diyagramları aracılığıyla bibliyometrik verileri  $tf*idf$  değerlerine dönüştürüp ilgililik ve erişim kolaylığı ile ilgili çıkarımlarda bulunmuştur. White'ın (2003, 2007a, 2007b, 2009, 2010,

2015) yararlandığı tamamen terim sıklığına dayanan vektör uzayı modelindeki yaklaşımın, konusal ilgililik ve o bilginin ele geçirilmesi için ne kadar çaba harcanmasıyla örtüştüğü görülmektedir (Allott, 2013).

*Ters belge sıklığı* yöntemi çekirdek terim (seed) ile birlikte geçen terimlerin ya da yazarların ilgililik puanlamasını yapmak için de kullanılmaktadır. White (2007a, 2007b, 2010) bir çekirdek terim ile birlikte birden fazla terimin (ya da yazarın) birlikte görülme sayısına göre sıralamalar oluşturmuştur. Bu dağılımlar klasik *tf\*idf* ağırlıklandırması sıralaması ile önemli farklılıklar göstermektedir. En baştaki terimler semantik olarak çekirdek ile belirgin olarak ilgiliyken daha aşağıdaki terimler daha genel özellikler taşımaktadır. Sık sık çekirdek ile birlikte görülenlerin semantik bağları daha az açıktır. Bu yüzden terimler kavramsal etkileri ve işleme çabalarında farklılık göstermektedir.

ERIC veri tabanından elde edilen verilerle yapılan bir çalışmada “bilgi gereksinimi” (information need) terimi çekirdek terim olarak ele alınmıştır (White, 2010). Sonuçlar *tf\*idf*e göre ağırlıklandırılıp sıralandığında üst sıradaki terimler “kullanıcı bilgi gereksinimi” ve “bilgi arama davranışı”dır. Bu terimlerin çekirdek terimle olan ilgililiklerini görmek için fazla çaba gerekmemektedir. Fakat daha aşağı sıralarda listelenen “toplum” ve “ilişki” gibi terimlere göre çekirdek terimle daha ilgilidirler. Alt sıralarda yer alan terimlerin çekirdek terim ile olan ilgisini görmek semantik uzaklıktan dolayı daha zordur. White aynı çalışmasında bu kez Katy Börner’i çekirdek yazar olarak ele almış, ortak atıf (*tf*) sayıları ve toplam atıf (*df*) sayılarını *tf\*idf* değerlerine dönüştürerek ilgililik teorisine göre yorumlamıştır. Ağırlıklandırılmış sıralamada veri görselleştirme, dijital kütüphaneler ve atıf analizi gibi terimler üstlerde yer almıştır. Bu terimler Börner’in araştırmaları ile yakından ilgilidir. Sıralamada aşağı doğru inildikçe diyagramlar ve karmaşıklık gibi daha genel ve göreceli olarak daha az ilgili terimler görülmektedir. Bu terimler tamamen ilgisiz olmamakla beraber genel olarak Börner’in çalışma alanı ile daha az ilgilidir. Çalışmada çekirdek olarak seçilen terim (bilgi gereksinimi) ve yazar (Katy Börner) ile ilgili bibliyometrik verilerden yararlanılarak oluşturulan asimetrik dağılımlarda ilgililiğe yeni bir yorum getirilmiştir.

Yüzeysel olarak farklı görünen ancak temelinde benzer özellikler gösteren birbirine yakın alanlardaki bilgilere erişim değerlidir. Pennant diyagramları yaklaşımını diğer yöntemlerden ayıran en önemli özellik bu alanlar arasındaki yakın ilişkileri ortaya çıkarmasıdır. Poetz ve Prügl (2010) birbirine yakın alanlardaki bilgilere erişim sağlamak için öne sürdükleri piramit yönteminde farklı uzmanlık alanlarından kişilerden tavsiyeler alındıkça orijinal soruna geri dönüp değiştirme ve geliştirme fırsatı oluştuğunu belirtmişlerdir. Bilimsel iletişim de bu şekilde ilerlemektedir. Bu durum dağıtık sorun çözme disiplini kapsamına girmektedir. Bu yüzden alanlar arasındaki gizli ilişkileri görmek her zaman önemlidir.

### **2.1.5. Atıf Dizinlerinde İlgililik Sıralaması**

Atıf dizinlerinde arama yapılan kaynağa benzer kaynakların araştırmacılara sunulması için ilgililik sıralamaları oluşturulmaktadır (Carevic ve Mayr, 2014). Sıralamalar çoğunlukla kaynakça benzerliğine bakılarak yapılmaktadır. İki çalışmanın kaynakçası ne kadar fazla ortak kaynak içeriyorsa bu iki çalışmanın aynı konuda olma olasılıkları o kadar yüksektir, dolayısıyla bu iki çalışma birbiriyle o kadar ilgilidir. Bu çalışma kapsamında kullanılan *Web of Science (WoS)* veri tabanında da ilgililik sıralaması kaynakçalara bakılarak yapılmaktadır (Web of Science, 2010). *WoS ilgili kayıtlar (related records)* özelliği kullanılarak arama yapılan makalenin kaynakçası ile veri tabanındaki diğer çalışmaların kaynakçalarına bakılarak hangi çalışmaların kaynakçalarının arama yapılan çalışmanın kaynakçasıyla daha çok örtüştüğü belirlenmektedir. İlgililik sıralaması, kaynakçası en çok örtüşen kaynaktan başlayarak listelenmektedir.

İlgililik sıralamasında kaynakça benzerliği dışında diğer yöntemler de kullanılmaktadır. Örneğin *Scopus*'ta benzerlik oranları belirlenirken ve sıralamalar oluşturulurken yazarlar ve anahtar kelimeler de dikkate alınmaktadır (Elsevier müşteri danışmanı, kişisel iletişim, 15 Aralık 2015).



### 2.1.5.1. Ortak Atıf Sayılarına Göre Oluşturulan İlgililik Sıralamaları

Pennant diyagramları yöntemi çekirdek terim (seed) ile birlikte geçen terimlerin ya da yazarların ilgililik puanlamasını hesaplamak için de kullanılmaktadır (White, 2007a, 2007b, 2010). Bu değer normalleştirilmiş  $tf$  ve  $idf$  değerlerinin çarpımından elde edilmektedir ( $1+\log(tf)*\log(N/df)$ ).

### 2.1.5.2. Kaynakça Benzerliği ve Ortak Atıf Sayılarına Göre Oluşturulan İlgililik Sıralamaları

Pennant yöntemi ile hesaplanan ilgililik sıralaması ile kaynakça benzerliğine göre hesaplanan ilgililik sıralaması büyük farklılık göstermektedir. Tablo 1'deki örnek verilere göre White çekirdek makalesinin kaynakçasında toplam 21 kaynak bulunmaktadır. Bu kaynaklardan dokuz tanesi Akbulut'un, iki tanesi de Cooper'ın makalesinin kaynakçasında yer almaktadır.

Tablo 1. White (2015) çekirdek makalesi için örnek veri seti

	Toplam Atıf Sayısı ( $df$ )	Ortak Atıf Sayısı ( $tf$ )	Kaynakçalardaki Ortak Kaynak Sayısı	$\log(1+tf)$	$\log(N/df)$	İlgililik ( $tf*idf$ )
<b>White, 2015 (Çekirdek)</b>	14	14	21	2,14	5,55	11,91
<b>Akbulut, 2014</b>	15	5	9	1,69	5,52	9,38
<b>Cooper, 2000</b>	6	4	2	1,60	5,92	9,48
<b>Garfield, 2012</b>	4	1	0	1,00	6,09	6,09

Tablo 2'de üç çalışmanın White çekirdek makalesi için oluşturulan sıralamalarına yer verilmektedir. WoS ilgili kayıtlar sıralamasına göre çekirdek makale ile en ilgili kaynak Akbulut, ikinci en ilgili kaynak ise Cooper'dır. Garfield'ın kaynakçasında çekirdek makalenin kaynakçasında olan herhangi bir kaynak bulunmadığı için ilgililik sıralamasına dahil olmamıştır.

Öte yandan pennant sıralamasına göre oluşturulan ilgililik sıralamasında Cooper'ın makalesi, Akbulut'unkine göre çekirdek makale ile daha ilgilidir. Akbulut'un toplam atıf sayısı daha yüksek olduğu için pennant diyagramları yöntemi hesaplaması ile baskılanmıştır. Cooper'ın makalesinin ise toplam atıf sayısı ve çekirdek makale ile birlikte anılma sayısı birbirine yakın olduğu için ağırlığı artmış ve çekirdek makale ile daha ilgili çıkmıştır.

Tablo 2. İlgililik sıralamaları

	İlgili Kayıtlar Sıralaması Sıra No	Pennant İlgililik Puanı ( $tf*idf$ )	Pennant İlgililik Sıralaması Sıra No
<b>White, 2015</b>	0	11,91	0
<b>Akbulut, 2014</b>	1	9,38	2
<b>Cooper, 2000</b>	2	9,48	1
<b>Garfield, 2012</b>	-	6,09	3

Bu çalışmada Maron ve Kuhns'un (1960) çekirdek makalesi için pennant diyagramları yöntemi ile ilgililik sıralaması hazırlanmış ve WoS'daki mevcut uygulamayla karşılaştırılmıştır.

## 2.2. LİTERATÜR DEĞERLENDİRMESİ

Bir önceki kavramsal arkaplan ile ilgili alt bölümde gerek bilgi erişim gerekse ilgililik ve pennant diyagramlarıyla ilgili pek çok çalışmaya atıf yapıldı. Bu alt bölümde büyük ölçüde yukarıda değinilmeyen kaynaklar değerlendirilmektedir.

Bilgi erişim sisteminin etkinliği kullanıcının sisteme girdiği sorguya karşılık ilgili bulacağı belgeleri çok ilgiliden daha az ilgiliye doğru sıralayabilmesi ile alakalıdır. Söz konusu ilgililik sıralamasının yapılabilmesi için benzerlik hesaplamaları yapılmaktadır. Bu aşamada terim ağırlıklandırması erişim performansını fark edilir derecede artırmaktadır (Chisholm ve Kolda, 1999; Hiemstra, 2000, Manning, Raghavan ve Schutze, 2008; Wu, Luk, Wong, ve Kwok, 2008). Bilgi erişim sistemlerindeki etkinliğin sağlanması amacıyla geliştirilen algoritmalar bibliyometri ile de ilişkilendirilmiştir. Bibliyometri ile bilgi erişimin ilişkilendirilmesi ile ilgili çalışmalardan ilki Small'un (1973) ortak atıf

analizini kullanarak iki belge arasındaki benzerliği hesapladığı çalışmasıdır. White ve Griffith (1981) ise Small'un iki belge ne kadar çok birlikte atıf almışsa bu belgeler o kadar çok birbirine benzer şeklinde açıkladığı benzerliği iki yazar arasında hesaplamış ve yazar ortak atıf analizini (author co-citation analysis) geliştirmişlerdir.

Bibliyometrik veriler arama motorları (örneğin, Google'ın PageRank algoritması) ve atıf dizinlerinde ilgililik sıralaması oluşturmak amacıyla sıklıkla kullanılmaktadır. Bibliyometrik verilere ek olarak makalelerin web üzerindeki etkisinin de ölçümüne yönelik olarak Altmetri (altmetric) istatistikleri (paylaşım sayısı, retweet sayısı vs.) kullanılmaktadır (Thomson Reuters, 2015)

Son yıllarda uluslararası konferanslarda (ECIR, CEUR, CHIIR, ISSI vb.) bibliyometri ve bilgi erişimin ilişkilendirildiği çok sayıda bildiri sunulmaktadır. Bu bildirilerde bibliyometrik hesaplamalar ve görselleştirme amacıyla kullanılan araçların bilgi erişimde kullanılması, bibliyometrik tavsiyeler ile klasik bilgi erişimin birleştirilmesi gibi konular tartışılmaktadır. Bibliyometri ile bilgi erişim arasındaki tamamlayıcı ilişki giderek daha yoğun ilgi görmektedir (Wolfram, 2000; Wolfram, 2015). ISSI2013 "Combining Bibliometrics and Information retrieval" çalıştayında sunulan bildiriler *Scientometrics* dergisinin özel sayısında yayımlanmış, bu özel sayının editoryalinde bibliyometri ve bilgi erişime ile ilgili temel kavramlara ve aralarındaki ilişkilere değinilmiştir (Mayr ve Scharnhorst, 2015).

Temeli ilgililik teorisi,  $tf*idf$  ve bibliyometriye dayanan pennant diyagramlarının etkileşimli uygulamasını içeren AuthorWeb sistemi Howard White, Xia Lin ve Jan Buzydlowski'nin bulunduğu bir araştırma ekibi tarafından geliştirilmiştir (White, 2015; White, Buzydlowski ve Lin, 2000). Çekirdek yazar ve ortak atıf alan yazarları görselleştirerek çekirdek yazarın entellektüel ortaklarını izlemeye olanak sağlayan AuthorWeb; Pathfinder ağları<sup>5</sup> (pathfinder Networks (PFNETs)), Kohonen özörgütlemeli haritalar<sup>6</sup> (self-organizing maps (SOMs)) ve

<sup>5</sup> PFNETs çift yazarlar arası güçlü bağlantıları ortaya çıkarmaktadır.

<sup>6</sup> Kohonen özörgütlemeli haritaların oluşturulmasında yazarlar arası uzaklık metaforları kullanılmaktadır.

pennant diyagramları olmak üzere üç tür ortak atıf haritası oluşturmaktadır (White, 2003). Sistem ortak atıfları anında haritalandırmakta ve bunu yaparken bibliyometrik veriler, görselleştirmeler ve belge erişimini birleştirmektedir (Buzydlowski, 2002; Lin, White ve Buzydlowski, 2003; White, Lin ve Buzydlowski, 2001).

Pennant diyagramlarında *terim sıklığı* değeri bir terimin kullanıcı üzerindeki bilişsel etkisini, *idf* değeri ise kullanıcının bu terimi anlamak için sarfettiği çabayı temsil etmektedir (White, 2007b, 2015). *Terim sıklığı* ile *ters belge sıklığı* değerinin çarpımı, bilişsel etki ile erişim kolaylığını çarpmaya benzetmektedir. Ancak yüksek ilgililiğin işleme çabasını azaltması gerektiği için yüksek *idf* değeri daha az çabayı öngörmektedir. Bu yüzden White (2007a, 2007b) işleme çabasını erişim kolaylığı olarak adlandırmıştır. Böylece, her ne kadar ilgililik teorisindeki gibi işleme çabasının bir ölçüsü olarak kalacaksa da, yüksek ilgililik değeri kolaylık, düşük ilgililik değeri ise zorluk anlamına kullanılmaktadır. İşleme kolaylığı, çekirdek terim ile diğer terimler arasındaki bağlantının ne kadar kolay anlaşılabilirliği ile ilgilidir.

White'ın pennant diyagramı çalışmaları bilgi erişimin bibliyometrik dağılımlar ile ilişkisi açısından oldukça ilgi çekicidir. İki bölüm halinde yayımlanan ve pennant diyagramlarının ayrıntılı bir şekilde incelendiği çalışmasında White (2007a, 2007b), Melville'in dünyaca ünlü romanı Beyaz Balina'nın (Moby Dick) ve Harter'ın ilgililik teorisini bilgin bilim ile ilişkilendirdiği çalışmasının literatüre etkilerini incelemiştir. Olle Persson'ın Price ödülüne layık görüldüğü sene hakkındaki bir anı kitabında ise Olle Persson and August Strindberg için diyagramlar oluşturmuştur (White, 2009). Söz konusu çalışmalarda pennant diyagramları yenilikçi uygulamalarında çalışmalar ve yazarlar arasında yeni ilişkilerin keşfedildiği ilginç bulgular yer almaktadır.

Tonta ve Özkan Çelik (2013) Türk matematikçi Cahit Arf'ın 1941 yılında yayımlanan *Arf invariantları* ile ilgili makalesi için pennant diyagramı hazırlamış *sosyal ağ analizi* (social network analysis) ve *tek yayın indeksi* (p-indeks) ile birlikte yorumlamışlardır. Böylece 1939-1966 yılları arasında üretim yapan (atıf dizinleri çıkmadan önce) ve yayınları atıf dizinlerinde neredeyse hiç atıf almamış

Arf'ın literatüre olan doğrudan ya da dolaylı etkisi geriye dönük olarak izlenebilmiştir.

Ortak atıf sıklığı iki çalışmanın benzerliği hakkında fikir verdiği için tavsiye sistemlerinde de tercih edilmektedir (Beel, Gipp, Langer ve Breitinger, 2015). Carevic ve Mayr (2014) ortak atıf değerleri kullanılarak oluşturulan pennant diyagramlarını *sowiport* dijital kütüphanesine (<http://sowiport.gesis.org>) entegre ederek tavsiye sistemi oluşturmuşlardır.

İlgililik sıralamaların araştırmacılara arama yaptıkları kaynağa benzer çalışmaların sundukları için tavsiye sistemlerindeki en önemli unsurdur. Bu açıdan da Carevic ve Mayr'ın (2014) pennant diyagramlarını entegre ettikleri *sowiport* dijital kütüphanesi uygulaması ilgililik ve tavsiye sistemleri ilişkisi açısından önemlidir.

*Referans Yılı Spektroskopisi* (Reference Publication Year Spectroscopy - RPYS) uygulaması ile bibliyometrik veriler kullanılarak oluşturulan etkileşimli grafikler sayesinde özgün makaleler, atıf klasikleri ve disiplinlerin dinamiklerini belirlenebilmektedir. Standart RPYS ve çoklu RPYS (multi-RPYS) olmak üzere iki çeşit RPYS analizi bulunmaktadır. Standart RPYS daha çok bir araştırma alanının özgün çalışmalarını belirlemek için kullanılmaktadır. Çoklu RPYS ise atıf alan çalışmaların atıf yapılan çalışmaya katkıları hakkında bilgi vermektedir. RPYS uygulaması aynı zamanda atıf oranlarındaki değişimleri de izlemeye olanak sağlamaktadır (Comins ve Leydesdorff, 2016a, 2016b). Çoklu RPYS atıf veren çalışmalar kümesini bölümlendirerek ve her yıl için standart RPYS'yi hesaplar. Oluşturulan ısı haritası (heat map) ile atıf verilen çalışmaların atıf kümesi üzerindeki etkileri görselleştirilmektedir.

## 3. BÖLÜM: YÖNTEM

### 3.1. GİRİŞ

Bu bölümde Maron ve Kuhns (1960) makalesine yapılan atıflar değerlendirilerek, makalenin daha sonraki dönemlerde yapılan diğer çalışmalara etkisinin ortaya çıkarılması için uygulanan yöntem, verilerin toplanması, analize uygun hale getirilmesi ve görselleştirilmesi ile ilgili bilgi verilmektedir.

### 3.2. MARON VE KUHNS'UN (1960) ÇEKİRDEK MAKALESİ

Maron ve Kuhns (1960) bir belgenin belirli bir sorguya karşılık ilgili bulunup bulunmayacağına dizincilerin bakış açısından yaklaşarak *olasılıksal (probabilistic) modeli* geliştirmişlerdir. Söz konusu model "Belgeler hangi özellikleri (örneğin hangi dizin terimlerini) içerirse belirli bir sorguya yararlı olur ya da sorguyu yöneltten kişi tarafından ilgili bulunur?" sorusunu adreslemeye yöneliktir. Formülasyon ise "Bir kullanıcı belirli bir terimi içeren sorguyla sisteme gelirse bu kullanıcının arama sonucu erişilen belirli bir belgeyi ilgili bulma olasılığı nedir?" şeklinde oluşturulmuştur.

İyi bir bilgi erişim sisteminin kullanıcının sorgusuna göre koleksiyondaki hangi belgelerin daha ilgili olduğunu öngörmesi ve bu belgeleri ilgililik olasılığı değerlerine göre sıralaması beklenmektedir. Olasılıksal modelde sıralama, bir belgenin sorguya karşılık ilgililiği ile o belgenin o terimi içermesi olasılığının çarpımı ile elde edilen bir değere bağlı olarak hesaplanmaktadır. Buradaki olasılık dizinlemeyi yapan kişi tarafından atanmaktadır. Fakat bu olasılık (varsa) önceki kullanımlara dayanarak da atanabilir. Böylece belgenin en azından öznel olarak ne kadar ilgili olduğu öngörülebilmektedir. Maron ve Kuhns bu teoriyi geliştirdiklerinde Internet ve web olmadığı için kullanıcıların erişilen belgelerin ne kadarını ilgili buldukları, ne kadarını bulmadıkları konusunda bilgi toplamak kolay değildi. Yazarlar bunun pratik olarak kolay olmayacağını ama kuramsal olarak bunun yapılabileceğini belirtmiş ve ödünç verme istatistiklerine dayanarak *popülerlik* katsayısını hesaplamayı önermişlerdir.

Belgelerle ilgili kullanım sıklığı vb. bilgiler toplanmış olsa bile yine de dizincilerin işi kolay değildir. Örneğin, ağırlıklandırma yöntemi kullanılan bir diğer model olan vektör uzayı modelinde ağırlıklar hesaplanırken bir terimin belgede geçme sıklığı, koleksiyonda geçme sıklığı, ters belge sıklığı vb. gibi değerler dikkate alınır. Dolayısıyla dizincinin bilişsel bir çabası söz konusu değildir. Oysa olasılıksal dizinlemede dizinleme yapan kişi gelebilecek bütün bilgi ihtiyaçlarına karşı, bir belgeye hangi terimler hangi olasılıkla atanırsa bu belge daha çok kullanılır ya da daha çok kişi tarafından ilgili bulunur soruları için zihinsel bir çaba harcamak zorundadır.

Maron ve Kuhns'un makalesi olasılıksal bilgi erişim hakkındaki ilk makaledir. Bunun yanı sıra makalede sorgu genişlemesi ve ilgili erişilen belgelerin genişlemesi kavramları ilk kez tanıtılmakta, önerilen arama stratejileri ve ampirik verilere dayanan sonuçlar tartışılmaktadır.

Bu çalışmada bilgi erişim alanında önemli bir yere sahip olan ve birçok erişim algoritmasını etkileyen Maron ve Kuhns'un (1960) olasılıksal erişim makalesinin etkisi pennant diyagramları ile yorumlanmıştır. Söz konusu çalışma için pennant diyagramları yöntemi kullanılarak ilgililik sıralamaları oluşturulmuştur.

### **3.3. VERİLERİN TOPLANMASI VE ANALİZE UYGUN HALE GETİRİLMESİ**

Araştırmada *Web of Science* veri tabanında dizinlenen ve Maron ve Kuhns (1960) çekirdek makalesine atıfta bulunan 294 çalışma ile bu çalışmalara ait 9607 (tekil sayı 4176'dır) referansın bilgileri ".txt" ve ".xls" formatında indirilmiş ve incelenmiştir.

WoS'un mevcut arayüzünde çekirdek makaleye atıf yapan 294 çalışmanın kaynakçasında yer alan çalışmaların listelenmesi (söz konusu 294 çalışma ile bağlantılarını koruyarak) mümkün değildir. Bu yüzden her çalışma kaynakçası ile birlikte ayrı ayrı indirilmiştir. Ardından bu veriler bir MS Excel dosyasında referansların hangi makaleye dahil olduğu bilgisi de eklenerek birleştirilmiştir.

Etkileşimli zaman çizelgesinde (time slider) makalelere yapılan atıfların ve bu makalelerin çekirdek makale ile birlikte aldığı atıfların yıllara göre dağılım

bilgileri kullanılmıştır. MS Excel programında yazılan makrolar ile yıllara göre ortak atıf sayıları ve toplam atıf sayıları hesaplanarak etkileşimli grafikler hazırlanmıştır.

Kaynakçasında Maron ve Kuhns (1960) bulunmamasına karşın başlığında, özetinde ve anahtar kelimelerinde *olasılıksal model, model 1, terim ağırlıklandırma* geçen kaynaklar da veri setine dahil edilmiştir. Söz konusu çalışmalara yapılan atıfların atıf dizinleri üzerinde ne ölçüde temsil edildiğini bulmak için çeşitli sorgu cümleleri ile ayrıntılı taramalar yapılmıştır (cited reference search, topic search vb.).

Bunun yanı sıra Maron ve Kuhns'un (1960) makalesi için WoS'un ilgili kayıtlar yöntemi ile oluşturulan ilgililik sıralaması ile kaynakça benzerliğine dayalı olarak oluşturulan ilgililik sıralamaları karşılaştırılmıştır. Maron ve Kuhns'un (1960) çekirdek makalesi ile ilgili kayıt sayısı pennant yöntemi kullanılarak 4176, WoS'un ilgili kayıtlar yöntemi kullanılarak ise 9803 olarak bulunmuştur.

İlgili kayıtlar yönteminde bir makale çekirdek makalenin kaynakçasındaki sadece bir kaynağı bile içerse bu makale ilgililik sıralamasına girmektedir. Pennant diyagramları yönteminde ise çekirdek makale ile ortak atıf sıklığına bakıldığı için ilgililik sıralamasına dahil olan kaynak sayısı daha azdır.

Pennant diyagramlarının oluşturulması için uygulanan işlemler Ek 1'de görselleştirilmiştir.

### 3.3.1. Çekirdek Makale ile Birlikte Geçiş Sıklığının (*tf*) Hesaplanması

İlk aşamada çekirdek makaleye atıf yapan çalışmalar kaynakça bilgileri ile birlikte WoS'tan ".txt" formatında indirilmiştir. Pennant diyagramının x ekseninde yer alacak çekirdek makale ile birlikte geçiş sıklığının (*tf*) yıllara göre hesaplanması için bir makro yazılmıştır.<sup>7</sup> Hesaplama için WoS'un standart veri dosyasında yer alan ve yıl bilgisini içeren PY (Publication Year) alanı ile referans bilgilerini içeren CR (Cited References) alanında yer alan bilgiler

<sup>7</sup> Çekirdek makale ile birlikte geçiş sıklığını hesaplamak için yazılan makroya erişim adresi: [http://www.mugeakbulut.com/YL\\_Tez/makrolar/tf\\_hesaplama\\_makrosu/tf\\_hesaplama.xlsm](http://www.mugeakbulut.com/YL_Tez/makrolar/tf_hesaplama_makrosu/tf_hesaplama.xlsm)



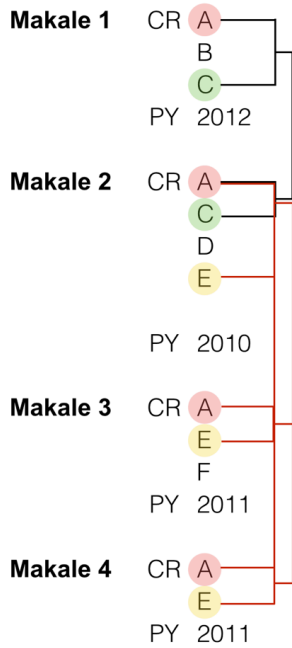
kullanılmıştır. Yazılan makro aracılığıyla CR alanındaki her bir satır yeni bir MS Excel sayfasına (tekrarlar temizlenerek) yazdırılmış ve söz konusu kaydın PY alanındaki yıl için atıf sayısı bir artırılarak yıllara göre dağılım hesaplanmıştır. Makronun üzerinde çalıştırıldığı dosya çekirdek makaleye atıf yapan çalışmalardan oluştuğu için her kaydın CR alanında mutlaka çekirdek makale geçmektedir. Dolayısıyla her bir referans için CR alanlarındaki geçiş sıklıkları aslında çekirdek makale ile birlikte anılma sıklıklarını vermektedir. İlgili kaydın yılına (PY) bakarak her bir referansın hangi yılda kaç kez çekirdek makale ile birlikte atıf aldığı hesaplanmıştır. Yıl aralığı çekirdek makalenin yayınlandığı yıl ile 2015 arasındadır.

Şekil 2, WoS standart veri dosyasının örnek gösterimidir. Şekilde dört makalenin yayın yılları (PY) ve referanslarına (CR) ilişkin bilgiler bulunmaktadır. WoS standart dosyada çok daha fazla alan etiketi bulunmaktadır<sup>8</sup> fakat araştırma kapsamında yapılacak hesaplama için bu iki alan yeterli olduğu için örnek de buna göre hazırlanmıştır.

Şekildeki dört makalenin de kaynakçasında A bulunduğundan, A bu örnek için çekirdek makaledir. Dolayısıyla herhangi bir referans veri setinde bir kere bile geçse A ile yani çekirdek makale ile birlikte görüldüğü anlamına gelmekte ve bu özelliğinden dolayı pennant diyagramında bir nokta olarak yerini almaktadır.

Makale 1'in kaynakçasında üç, Makale 2'nin kaynakçasında ise dört çalışma bulunmaktadır. A ve C, biri Makale 1'den diğeri de Makale 2'den olmak üzere iki ortak atıf almışlardır. Yani A ve C'nin ortak atıf sayısı ikidir. Aynı şekilde A ve E'nin ortak atıf sayısı ise üçtür. Başka bir ifadeyle, Makale 2, Makale 3 ve Makale 4'ün kaynakçasında A ve E yer almaktadır.

<sup>8</sup> WoS'da 60 farklı alan etiketi bulunmaktadır. Bu etiketlerden tanımlanmış olanlar her bir makalenin altında sıralanmaktadır. Tüm etiketler ve anlamları için bkz. [http://images.webofknowledge.com/WOKRS55B6/help/WOS/hs\\_wos\\_fieldtags.html](http://images.webofknowledge.com/WOKRS55B6/help/WOS/hs_wos_fieldtags.html)



Tablo 3. Sıklık verilerinin yıllara göre dağılımı

Referans	2010	2011	2012
A	1	2	1
B	0	0	1
C	1	0	1
D	1	0	0
E	0	2	0
F	0	1	0

Şekil 2. WoS standart veri dosyası örnek gösterimi

Hem bibliyografik eşleşme hem de ortak atıf analizi için en az iki kaynağa ihtiyaç vardır. Şekil 2'de A ve C için ortak atıf söz konusudur. Bu durumun oluşması için gerekli iki kaynak Makale 1 ve Makale 2'dir. Öte yandan A ve B için ortak atıf durumu yoktur çünkü sadece bir kaynakta A ve B'ye atıf yapılmıştır. Derlemde A ve B'ye atıf yapan ikinci bir kaynak daha bulunmamaktadır, dolayısıyla burada Makale 1'de A ve B'ye birlikte atıf (co-cited) yapılmıştır.

Çalışmada kullanılan veri setinin yapısından dolayı her bir referans için hesaplanan sıklık verileri (*tf*) aslında çekirdek makale ve ilgili referansa birlikte atıf yapılma sayılarını, yani birlikte görünme sayısını vermektedir. Şekil 2'deki örnek için hesaplanan sıklık verilerinin yıllara göre dağılımı Tablo 3'te verilmektedir.

### 3.3.2. Toplam Atıfların Yıllara Göre Dağılımının Hesaplanması

İkinci aşamada *idf* bilgileri hesaplanmıştır. *idf* değerleri çekirdek makalelere atıf yapan çalışmaların referans bilgileri ve onlara yapılan atıfların bilgilerinin yıllara

göre dağılımı ile ilgilidir. İkinci derece atıf bilgisi (çekirdek çalışmalara atıf yapan çalışmalara yapılan atıflar) mevcut sistem üzerinde ayrı bir yerde tutulmakta ve bu veriler otomatik olarak elde edilememektedir.<sup>9</sup> *Idf* değerlerinin sağlanması için her bir kaynağa gidilerek referansları (cited references) işaretli listeye (marked list) aktarılmış ve bu listeden atıf raporları (citation reports) oluşturulmuştur.

Buradaki bir diğer sorun *WoS*'daki dış kaynak<sup>10</sup> verilerinden kaynaklanmaktadır. Bu kayıtlar işaretli listeye atılamadığı için yıllara göre dağılım bilgileri de bu aşamada elde edilememiştir. Fakat atıfların önemli bir kısmı da bu listelenmeyen dergilerde yayımlanan dergi ve kitaplardan gelmektedir. *WoS*, "Cited References" aramasında listelenen kısaltılmış, tıklanamayan atıfların da varsa yıl bilgilerini listelemektedir. Bunun için örneğin, vektör uzayı modeli çekirdek makale için tüm "Cited References" sonuçları dökülüp tıklanabilir olanları ayıklanmış, daha sonra içinde rakam geçen satırlar işaretlenerek gerçekten yıl bilgisi olanlar (cilt ve sayı bilgileri de olabilir) gözle kontrol edilmiştir.

### 3.3.3. Standardizasyon

Atıf dizinlerinde standardizasyon sorunları bulunmaktadır. Bu sorunlar genelde yazarların farklı biçimde atıf vermesinden ya da elci (manuel) sistemden kaynaklanmaktadır. Bu tür durumlarda *WoS* farklı biçimde atıf verilen çalışmaları da listelemektedir. Bu da bibliyometrik çalışmalarda hatalı sonuçlar elde edilmesine neden olmaktadır. Çalışma kapsamında indirilen veriler standardizasyon sorununun en aza indirilebilmesi için tekbiçim hale getirilmiştir. Bu amaçla bir benzerlik makrosu<sup>11</sup> hazırlanmıştır. Birleştirme işleminin doğru yapılabilmesi için makrodaki benzerlik oranı yüksek tutulmuş (%80) ve makro çıktısı birleştirme yapılmadan önce gözle kontrol edilmiştir.

<sup>9</sup> Sadece DOI bilgisi olan kayıtlar için API'lar kullanılarak indirme yapılabilmektedir.

<sup>10</sup> İlgili kaynağa atıf yapan kaynak *WoS*'da dizinlenmemişse dış kaynak olarak adlandırılmaktadır.

<sup>11</sup> Hazırlanan birleştirme makrosuna [http://www.mugeakbulut.com/YL\\_Tez/makrolar/birlestirme\\_makrosu/all\\_versions/benzerlik\\_algoritmali/](http://www.mugeakbulut.com/YL_Tez/makrolar/birlestirme_makrosu/all_versions/benzerlik_algoritmali/) adresinden erişilebilir.

Yazar isimleri ile ilgili standardizasyon sorunlarına örnek vermek gerekirse: İstatistiksel ağırlıklandırma yöntemini ortaya atan Karen Sparck Jones'a çok farklı şekillerde atıf verilmektedir (SPARCKJONES K, Sparck-Jones K, Sparck Jones SK, SPARCK-JONES K, Sparck Jones Karen, SPARCK JK, Sparck J K, Jones SK, Jones Sparck, Jones KS gibi). Bu girişler farklı kişiler gibi işlem görmektedir. Dolayısıyla Sparck Jones'a yapılan atıflar sistemde olduğundan çok daha az gözükmektedir. Aynı kişiye ait olan bu tür girişler benzerlik algoritması ile belirlenerek tekbiçim hale getirilmiştir.

Bunun dışında aynı kaynağa yapılan atıfların da farklı kaynaklara yapılmış gibi gözüktüğü durumlar olmuştur. Örneğin bir kitaba (SALTON G, 1975, DYNAMIC INFORMATION) yapılan atıf ile aynı kitabın belli bir bölümüne (SALTON G, 1975, DYNAMIC INFORMATION, CH6) yapılan atıf iki farklı atıf olarak gözükmektedir. Buna benzer durumlar da benzerlik algoritması ile belirlenerek standart hale getirilmiştir.

#### **3.3.4. Ortak Atıf ve Toplam Atıf Sayılarının Birleştirilmesi ve Etkileşimli Zaman Çizelgesinin Hazırlanması**

*Tf* ve *idf* değerleri toplanıp düzenlendikten ve atıf sayılarının yıllara göre dağılımı hesaplandıktan ve birikimli hale getirildikten sonra veriler etkileşimli çizelge için uygun hale dönüştürülmüştür. Bu aşamada *tf* değerlerinin yıllara göre dağılımı makrosu sonucu ile *idf* değerleri için hazırlanan atıf raporları sonuçları birleştirilmiştir.

Şekil 3, ortak atıf saydırma makrosu çıktısı örneğidir. Burada ilk sütunda yayın bilgileri yer almaktadır. Bu format *WoS*'un standart veri dosyasındaki CR alanı formatıdır. Sadece ilk yazarın soyadı ve adının ilk harfi, yayın yılı, dergi adı (kısaltılmış halde) ve varsa sayı, sayfa, DOI gibi bilgiler tek bir sütunda yer almaktadır. Diğer sütunlarda ise çekirdek makalenin yayın yılından 2015 yılına kadar çekirdek makale ile birlikte görünme sıklıkları yer almaktadır.

	A	B	C	D	E	F	G	H	I	J	K
1	TF_MACR_CR	TF_1960	TF_1961	TF_1962	TF_1963	TF_1964	TF_1965	TF_1966	TF_1967	TF_1968	TF_1969
2	KUIPERS JW, 1957, AM DOC, V8, P246, DOI 10.1002/asi.5090080403	1	0	0	0	0	0	0	0	0	0
3	MARON ME, 1961, J ACM, V8, P404, DOI 10.1145/321075.321084	0	0	2	0	0	3	1	0	0	2
4	SWANSON DR, 1960, SCIENCE, V132, P1099, DOI 10.1126/science.132.3434.1 099	1	0	1	1	2	0	0	0	0	0
5	Zhang C, 2010, IEEE NETWORK, V24, P13	0	0	0	0	0	0	0	0	0	0

Şekil 3. Ortak atıf saydırma makrosu sonucu

Şekil 4'teki dosyayı elde edebilmek için çekirdek makaleye ait tüm kaynakçalar işaretli listeye kaydedilerek bu listeden atıf raporları oluşturulmuştur. *Idf* değerlerinin yer aldığı raporda her bir kayıt için farklı sütunlarda başlık, tüm yazarlar, kaynak, yayın yılı, başlangıç ve bitiş sayfaları, konferans bilgisi vb. ayrıntılı bilgiler bulunmaktadır.

	A	B	C	D	E	F	G	H	I	J	K
1	IDF_Title	IDF_Authors	IDF_Source Title	IDF_Publication Year	IDF_Volume	IDF_Issue	IDF_Beginning Page	IDF_Ending Page	IDF_1960	IDF_1961	IDF_1962
2	A MINICARD SYSTEM FOR DOCUMENTARY INFORMATION	Kuipers, JW; Tyler, AW; Myers, WL	AMERICAN DOCUMENTATION	1957	8	4	246	268	2	0	2
3	AUTOMATIC INDEXING - EXPERIMENTAL INQUIRY	Maron, ME	JOURNAL OF THE ACM	1961	8	3	404	412	0	0	2
4	SEARCHING NATURAL LANGUAGE TEXT BY COMPUTER	Swanson, DR	SCIENCE	1960	132	3434	1099	1104	1	1	5
5	Privacy and Security for Online Social Networks: Challenges and Opportunities	Zhang, Chi; Sun, Jinyuan; Zhu, Xiaoyan; Fang, Yuguang	IEEE NETWORK	2010	24	4	13	18	0	0	0

Şekil 4. Atıf raporu

Bu kayıtların birleştirilmesi için bir makro daha yazılmıştır.<sup>12</sup> Eğer ilk dosyadaki CR alanındaki kaydın ilk kelimesi ve yazarın adının baş harfi ile virgülden sonra yer alan yayın yılı atıf raporundaki ilgili sütunlarda geçiyorsa son aşama olarak dergi adına bakılmıştır. CR alanındaki kayıta dergi adı kısaltılmış biçimde, ikinci dosyadaki formatta ise uzun hali ile verilmektedir. Bu yüzden dergi adı kısaltma listesi<sup>13</sup> (Journal title abbreviations list) ve dergilerin tam adları makroya eklenerek yazar adı ile yayın yılı aynı olan kayıtlar için bu listenin taranması sağlanmıştır.

<sup>12</sup> *Tf* ve *idf* değerlerinin standart bir biçime dönüştürülmesi için yazılan makroya [http://www.mugeakbulut.com/YL\\_Tez/makrolar/birlestirme\\_makrosu/](http://www.mugeakbulut.com/YL_Tez/makrolar/birlestirme_makrosu/) adresinden erişilebilir.

<sup>13</sup> [http://images.webofknowledge.com/WOK46/help/WOS/A\\_abrvjt.html](http://images.webofknowledge.com/WOK46/help/WOS/A_abrvjt.html)

Söz konusu tanımlayıcıların üçünün de örtüştüğü kayıtlar için eşleşme sağlanmış olarak kabul edilmiş ve bu kayıtlar yeni bir sayfada yıllara göre *tf* ve *idf* bilgileri ile birlikte yazdırılmıştır. Eşleşmeyen kayıtlar ise (dizinleme hatası olanlar, dış kaynak olanlar, bir kısaltmanın birden fazla dergiye verilmesi nedeniyle eşleşmeyen kayıtlar vd.) WoS'da tekrar taranabilmesi için ayrı bir sayfaya yazdırılmıştır. Tüm veriler tek bir dosyada toplandıktan sonra zaman çizelgesi hazırlanması için gereken formata çevrilmiştir.

Atıf veri tabanlarında bazı alanlar elci sistem ile dizinlenmektedir. Bu durumlar tutarsızlıkla ilgili sorunların ortaya çıkmasına neden olmaktadır (Galvez ve Moya-Anegón, 2006). Toplanan verilerde atıf veri tabanlarının içeriklerinde yaşanan artıştan dolayı tutarsızlık olan kayıtlar temizlenip yeniden düzenlenmiştir.

### 3.3.5. Pennant Diyagramlarının Oluşturulması

Çekirdek makaleye atıfta bulunan kaynakların kaynakçasındaki her bir çalışma pennant diyagramında bir nokta olma potansiyeline sahiptir. Eğer söz konusu çalışma çekirdek makale ile birlikte bir kere bile anıldıysa pennant diyagramında bir nokta olarak yerini almaktadır.

Pennant diyagramlarında normalleştirilmiş değerlerin hesaplanması için kullanılan formüller aşağıdaki gibidir.

$$x \text{ akseni (bilişsel etki)} = 1 + \text{Log}(tf) \quad (1)$$

$$y \text{ akseni (işleme kolaylığı)} = \text{Log}(N/df) \quad (2)$$

Formüllerdeki *df* ortak atıf alan yazarların/çalışmaların toplam atıf sayısı, *tf* bir yazarın/çalışmanın çekirdek yazar/çalışma ile birlikte aldığı atıf sayısıdır. *N* değeri ise WoS veri tabanındaki toplam yazar/çalışma sayısıdır.

Tf ve df değerlerinde birbirine yakın olan çalışmaların yazarlarının çekirdek makalenin yazarı ile aynı konuda çalışmış ve aynı dönemlerde yaşamış

oldukları kabul edilir. Makalelerin çekirdek makale ile birlikte atıf alma sıklıkları aşağıda yer alan ağırlık ( $tf*idf$ ) formülüne (Manning ve Schütze, 2000, s. 542) göre hesaplanmış ve ilgililikleri belirlenmiştir.

$$tf*idf = \text{ilgililik} = [1 + \log(tf)] * [\log(N/df)] \quad (3)$$

Kolayca anlaşılabilceği gibi, ilgililik formülü (3), yukarıda tanımlanan bilişsel etki formülüyle (1) işleme kolaylığı formülünün (2) çarpımından oluşmaktadır. Çekirdek makale ile sadece bir kez birlikte anılan çalışmalar için ilgili değeri formüle yerleştirdiğimizde sonuç sıfır olur ( $\log(1) = 0$ ). Dolayısıyla bu tip çalışmalar y eksenine yapışacakları için örüntüyü gözlemek zordur. Bu durumu engellemek için bilişsel etki değeri hesaplanırken tüm değerlere sabit bir değer (1) eklenmiştir.

Maron ve Kuhns için oluşturulan pennant diyagramı 4177 noktadan oluşmaktadır (çekirdek makaleye 1960 ile 2015 yılları arasında atıf yapan 4176 çalışma artı çekirdek makale).<sup>14</sup> Diyagramda çok sayıda çalışma yer aldığından, ayrıntılı yorumlama yapabilmek için bu çalışmalardan kolayda örneklem (convenience sampling) tekniği kullanılarak çekirdek makale hariç 90 çalışmadan oluşan bir örneklem seçilmiştir. Diyagramdaki noktalar işleme kolaylığı ve bilişsel etki değerleri dikkate alınarak 10 alana ayrılmış ve her alandan (varsa) 12'şer çalışma seçilerek ayrıntılı içerik analizi yapılmıştır (çalışma sayısı 12'den az olan alanlardaki tüm çalışmalar örnekleme alınmıştır).

Çalışma kapsamında statik pennant diyagramlarını oluşturmak için *DeltaGraph* 7 (<http://redrocksw.com/>) programı kullanılmıştır.

### 3.3.6. Etkileşimli Zaman Çizelgesinin Oluşturulması

Etkileşimli zaman çizelgesi programının kodları için Google'ın API ve kaynak kodu paylaşım platformu Google Geliştirici'den (<https://developers.google.com/web/>) yararlanılmıştır. Veriler belli bir formatta düzenlendikten sonra hesaplamaların (logaritma, sektör vb.) yapıldığı bir veri

<sup>14</sup> Pennant diyagramının oluşturulması için kullanılan veri seti için bkz. <https://goo.gl/giWUgG>

tabanı hazırlanmıştır. Google geliştirici platformunda geliştirilen çizelge ile kolaylıkla uyum sağlayabilmesi için veri tabanı Google hesap tablosunda (spreadsheet) oluşturulmuştur.

### **3.3.7. Kaynakça Benzerliğine Dayalı İlgililik Sıralamasının Oluşturulması**

Pennant diyagramları yaklaşımında ortak atıf ve toplam atıf sayılarına bakılarak yapılan  $tf*idf$  ağırlıklandırma hesaplaması yöntemiyle elde edilen ilgililik sıralaması oluşturulmuştur. *MS Excel* programında yazılan makrolar ile ortak atıf sayıları ve toplam atıf sayıları hesaplanarak 4176 çalışmanın ilgililik sıralaması hazırlanmıştır. *WoS* veri tabanındaki ilgili kayıtlar özelliği ile kaynakçadaki örtüşmeye dayalı ilgililik sıralaması arasındaki benzerlik test edilmiştir. Listedeki çalışmalardan hangilerinin çekirdek makale ile ilgili, hangilerinin daha az ilgili olduğuna araştırmacı karar vermiştir. *WoS* ilgili kayıtlar özelliği ile oluşturulan ilgililik sıralamasının elde edilebilmesi için Maron ve Kuhns'un (1960) çekirdek makalesi için arama yapılmıştır. İlgili kayıtlar özelliği ile sıralanan kaynaklar işaretli listeye eklenmiş ve daha sonra .txt formatında dışa aktarılmıştır.



## 4. BÖLÜM: BULGULAR VE DEĞERLENDİRME

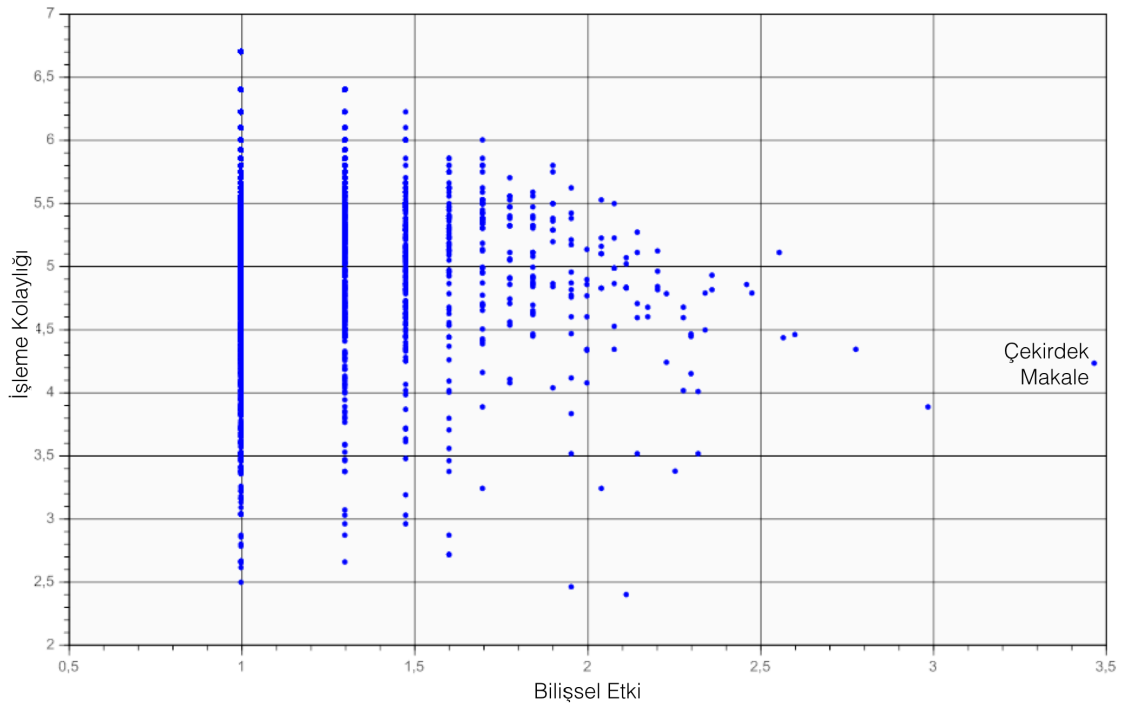
### 4.1. GİRİŞ

Çalışmanın bu kısmında birinci bölümde tanımlanan araştırma sorularını cevaplamak ve hipotezleri test etmek amacıyla Maron ve Kuhns (1960) çekirdek makalesi için oluşturulan pennant diyagramları ve ilgililik sıralamalarına ilişkin bulgular sunulmuştur. Öncelikle çekirdek makale ile en az bir kere birlikte anılmış 4176 çalışma üzerinden kolayda örneklem yöntemi ile seçilen 90 çalışmayı (çekirdek makale hariç) içeren pennant diyagramı ayrıntılı olarak incelenmiştir. Daha sonra WoS'un halihazırda kullandığı *kaynakça benzerliğine dayalı* ilgililik sıralamasına alternatif olarak çekirdek makalenin kaynakçası dışında diğer araştırmacıların atıflarının da hesaplama dahil edildiği *pennant diyagramı yöntemi* ile ilgililik sıralaması oluşturulmuştur.

### 4.2. MARON VE KUHN'S'UN (1960) ÇEKİRDEK MAKALESİ İÇİN PENNANT DİYAGRAMLARI

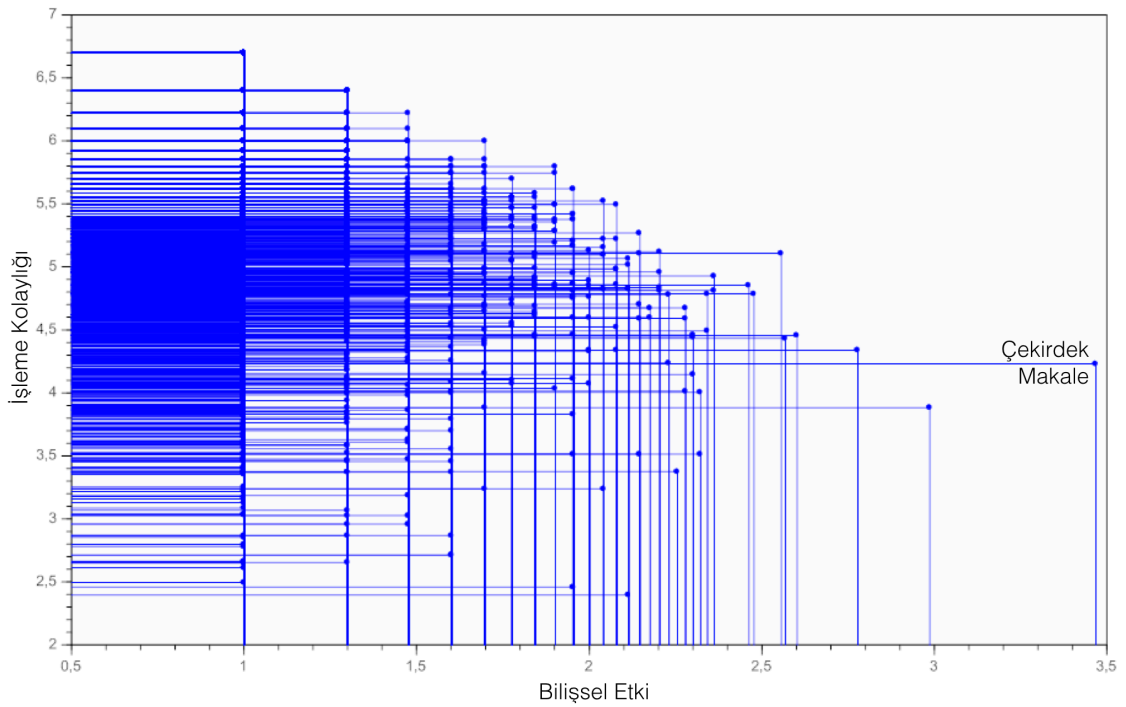
Çalışmada Maron ve Kuhns'un (1960) çekirdek makalesinin literatürdeki diğer çalışmaları nasıl etkilediği her bir yazarın çekirdek yazarlar (seed) ile ortak atıfları ve toplam atıflarının logaritmaları alınarak oluşturulmuş, bu etki pennant diyagramları aracılığıyla görselleştirilmiştir. Böylece söz konusu makalede savunulan olasılıksal modelin literatürde hangi yapıların/modellerin oluşmasında etkili olduğu ortaya çıkarılmıştır.

Bilgi erişim literatüründeki en önemli çalışmalardan birisi olan Maron ve Kuhns (1960) için oluşturulan pennant diyagramında 4176 çalışma (bkz. Şekil 5) bulunmaktadır. Bu çalışmalar çekirdek makaleye atıfta bulunan çalışmaların kaynakçasında yer alan tüm çalışmalardan oluşmaktadır. X eksenini (ortak atıf) çekirdek makaleyle birlikte atıf yapılan çalışmaların bilişsel etkisini, y eksenini (toplam atıf) ise bu çalışmalara erişim kolaylığını temsil etmektedir.



Şekil 5. Maron ve Kuhns (1960) pennant diyagramı

Şekil 6, diyagramdaki çalışmaların dağılımının yoğunluğunu göstermektedir. Bu gösterim için diyagramdaki her bir noktadan  $x$  ve  $y$  eksenlerine çizgiler çizilmiştir. Daha önce de belirtildiği gibi  $x$  eksenindeki değerler ortak atıf verilerinden oluşmaktadır.  $X$  ekseninde sağa doğru ilerledikçe çekirdek makale ile ortak atıf sayısı artmakta, diyagramın sağ yarısında çekirdek makale ile ortak atıf sayısı ve dolayısıyla bilişsel etkisi nispeten daha yüksek olan çalışmalar yer almaktadır. Kolayca görülebileceği gibi bu tür çalışmaların sayısı azdır. Benzeri bir biçimde,  $y$  ekseninde yukarı doğru çıkıldıkça diyagramın üst yarısında toplam atıf sayısı yüksek ve dolayısıyla daha kolay erişilen çalışmalar yer almaktadır.



Şekil 6. Maron ve Kuhns (1960) pennant diyagramı yoğunluk gösterimi

Tablo 4 ise söz konusu çalışmalar (4176 adet) için normalleştirilmiş *tf* ve *idf* değerlerinin dağılım özellikleri hakkında bilgi vermektedir. Tablo pennant diyagramına (bkz. Şekil 6) göre yorumlanacak olursa; satırlar y eksenindeki işleme kolaylığını, sütunlar ise x eksenindeki bilişsel etkiyi temsil etmektedir. Örneğin işleme kolaylığı değeri altı ve bilişsel etki değeri bir olan alan, *6-1 alanı* olarak adlandırılmıştır.<sup>15</sup> Bu alanda 646 tane çalışma bulunmaktadır. İşleme kolaylığı 4 bilişsel etkisi 3 olan tek çalışma çekirdek makaledir. En fazla çalışmaların ise 5-1 alanında bulunmaktadır (1806 adet). Bu alandaki çalışmaların toplam atıf ve ortak atıf değerleri çok düşüktür.

Diyagrama dahil olan çalışma sayısı nispeten fazla olduğundan (bkz. Şekil 6) çalışmaların içerik analizini yapmak için kolayda örneklem seçme yoluna gidilmiştir. Örneklem seçerken eğer bir alanda 12'den az çalışma varsa (4-1, 3-2 ve 2-2 alanları) bu çalışmaların hepsi diyagrama dahil edilmiştir, çalışma sayısı daha fazla ise bu alanlara dahil olan kaynaklar 12'şer tanesi alınmıştır (örneklemedeki toplam çalışma sayısı 90'dır). Bu sayı o alana dahil olan

<sup>15</sup> Okumayı kolaylaştırmak amacı ile alanlar; tablo, pennant diyagramları ve Ek 2 için renklendirilmiştir.

çalışmalar hakkında genel bir fikre sahip olmak için yeterlidir (White 2007a, s. 545). Örnekleme dayanan pennant diyagramının içerik analizi saçılım grafiğinin farklı bölümlerine dağılan çalışmaların hangi özelliklerde olduğunu anlamaya yardımcı olmaktadır.<sup>16</sup> Her ne kadar bu içerik analizi Maron ve Kuhns'un (1960) çekirdek makalesi için hazırlanmış olsa da söz konusu dağılım pennant diyagramları için tipiktir. Pennant diyagramındaki 90 çalışma hakkında ayrıntılı bilgi tablo Ek 2'dedir.

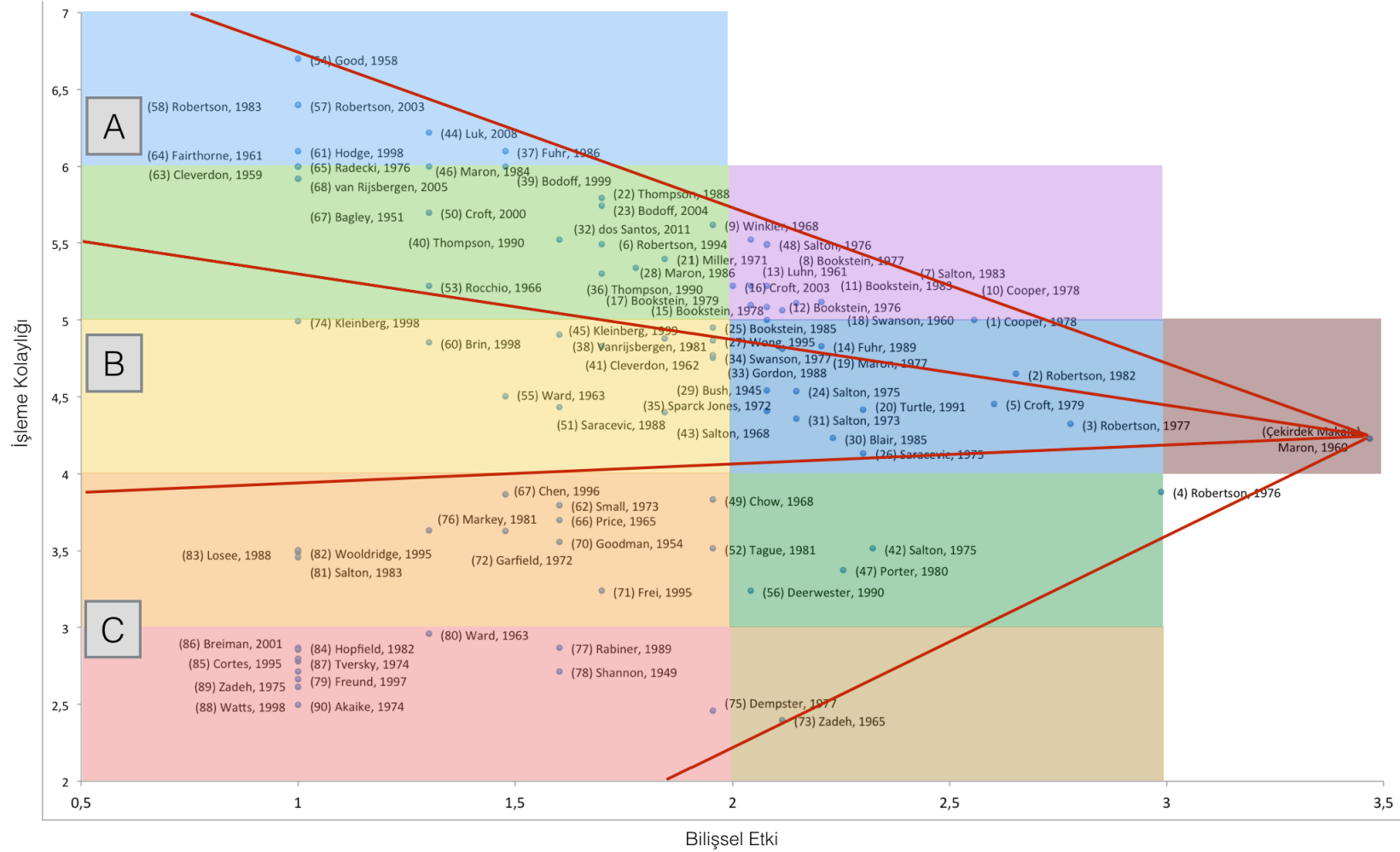
Tablo 4. Pennant diyagramındaki çalışmaların dağılımı

İşleme Kolaylığı	Bilişsel Etki			
		1	2	3
6		646		
5		1806	14	
4		1491	42	1
3		155	5	
2		16	1	
Toplam		4114	62	1

#### 4.2.1. Pennant Diyagramlarının İlgililik Açısından Yorumlanması

Pennant diyagramı yorumlanırken öncelikle çekirdekte yer alan yazar ya da çalışmalardan yola çıkılarak diyagramdaki noktalar üç gruba ayrılmaktadır (bkz. Şekil 7). Şekil 7'de isimlerin ya da konuların başındaki parantez içindeki sayı aynı zamanda ilgili kaydın pennant diyagramı yöntemi ile oluşturulan sıralamadaki *sıra numarasını* temsil etmektedir. Diyagramdaki çalışmalar hakkında ayrıntılı bilgi için ilgili sıra numarası takip edilerek Ek 2 ve Ek 3 incelenebilir.

<sup>16</sup> Pennant diyagramının (örneklem) etkileşimli versiyonu için bkz. <https://goo.gl/nW0Ik3>

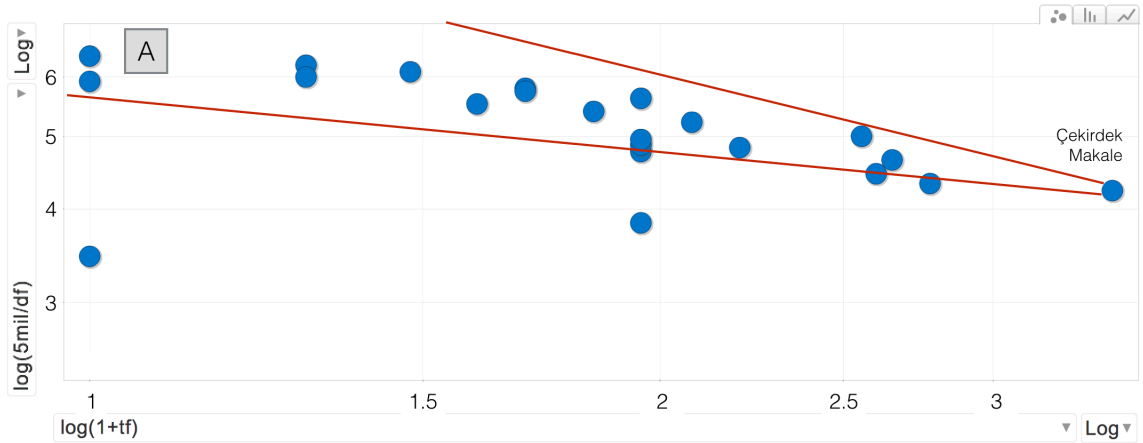


Şekil 7. Maron ve Kuhns (1960) pennant diyagramı (yazara göre)

Not: Parantez içindeki rakamlar ilgili kayıdın pennant diyagramı yöntemi ile oluşturulan sıralamadaki sıra numarasını temsil etmektedir. Bu kaynakların tam künyeleri Ek 2'de verilmektedir.

Çalışmaların pennant diyagramındaki konumları çekirdek makale ile ilgililikleri (konu ve yazarlar açısından) hakkında bilgi vermektedir. İşleme kolaylığı (y eksenini) açısından bakıldığında, A sektöründeki çalışmaların çekirdek makale ile ilgilerini görmek kolaydır, fakat B ve C sektörlerine doğru ilerledikçe aradaki ilgiyi görmek zorlaşır. Hatta C sektöründeki çoğu çalışmanın çekirdek makale ile ilgililiği ancak bu alanda uzman kişiler tarafından farkedilebilmektedir (White, 2015). Örneğin, “olasılık”, “ağırlıklandırma” veya “bilgi erişim” terimlerini *olasılıksal model* ile bağlantılandırmak nispeten kolaydır, fakat “PageRank” veya “rastgele orman” (random forest) kelimelerini bağlantılandırmak daha zordur. Bu durum literatürde konusal ilgililik (topical relevance) olarak adlandırılmaktadır. (bkz. konularına göre pennant diyagramı: <https://goo.gl/Iyooz3>)

Konusal ilgililik (topical relevance) açısından incelendiğinde, başlığında “olasılıksal” terimi geçen çalışmaların %81’inin A sektöründe olduğu görülmektedir. Bu sektördeki çalışmaların çekirdek makale ile ilgileri nispeten daha güçlüdür ve bu çalışmalar spesifik olarak çekirdek çalışmayla daha ilgilidir (bkz. Şekil 8).



Şekil 8. Başlığında “probabilistic” geçen çalışmalar (Etkileşimli diyagram için bkz. <https://goo.gl/h4RmpD>).

B sektöründeki çalışmalar başlıksal ilgililik içermekten ziyade daha derin ilgililik özelliği taşırlar. Başlık açıkça farklı olsa da içerik olarak benzerlikleri vardır. Genel bir ifadeyle B sektöründeki çalışmaların çekirdek çalışmada bahsedilen

konu veya terimlerle ilgili tartışmaları içermesi öngörülmektedir. Bu çalışmalar genellikle çekirdek makalede ortaya atılan modelin uygulamaları şeklindedir. Örneğin, Salton ve Lesk (1968) ile Hersh ve Voorhees'in (2009) ampirik verilere dayanan çalışmaları B sektöründe bulunmaktadır. Bu tür uygulamalı çalışmalarda Maron ve Kuhns'a (1960) ortak atıf olması beklenen bir durumdur. Kısacası, B grubundaki çalışmalar A grubundakilere göre daha derin ve başlıksal olmayan türde ilgililiğe sahiptir. C sektöründe ise çekirdek makale ile ilgililiği daha az açık olan başlıklar yer almaktadır. Bu çalışmalar A ve B sektöründekilere göre çoğunlukla daha geneldir (bkz. Şekil 7).

Örneğin, A sektöründe bulunan *Probabilistic Retrieval Models* (Maron, 1984) başlıklı çalışma çekirdek makale ile oldukça ilgilidir ama C sektöründeki *A Mathematical Theory of Communication* (Shannon, 1948) ya da *Support-Vector Networks* (Cortes ve Vapnik, 1995) çalışmalarının olasılıksal model ve ilgililik alanı dışında da kullanımları vardır (bkz. Şekil 7). Dolayısıyla bu çalışmalar ek bağlamlardan da atıf aldıkları için toplam atıf sayıları artmıştır.

Yazarlarla ilgili olarak, A sektöründe yer alan yazarların çekirdek makalenin yazar(lar)ının ardılları, B sektöründe yer alan yazarların çekirdek makale yazar(lar)ının akranları, C grubundakilerin ise çekirdekte yer alan yazar(lar)ın öncülleri olması beklenmektedir (bkz. Şekil 7). A sektöründeki yazarların çekirdek yazarla ilgisinin net olması beklenirken C kesimine doğru inildikçe bu bağlantıların kurulabilmesi için alandan uzman görüşüne ihtiyaç duyulmaktadır. Yazarların yaşlarının da bu ilişkiye göre yukarıdan aşağıya artması beklenmektedir. Ayrıca diyagram üzerinde yer alan yazarların, soldan sağa doğru çekirdek yazara yaklaşırken bilişsel olarak çekirdek yazar ile daha fazla ilgili olması öngörülmektedir (Sperber ve Wilson, 1995; Tonta ve Özkan Çelik, 2013; White, 2007a, 2007b, 2009, 2010, 2011, 2015; White ve Mayr, 2013). Çekirdek makalenin yazarlarından (Maron ve Kuhns) önce veya sonra özel olarak olasılıksal modeller konusunda ve yakın konularda çalışan yazarlar Cooper, Robertson, Croft, Sparck Jones ve Salton'dır

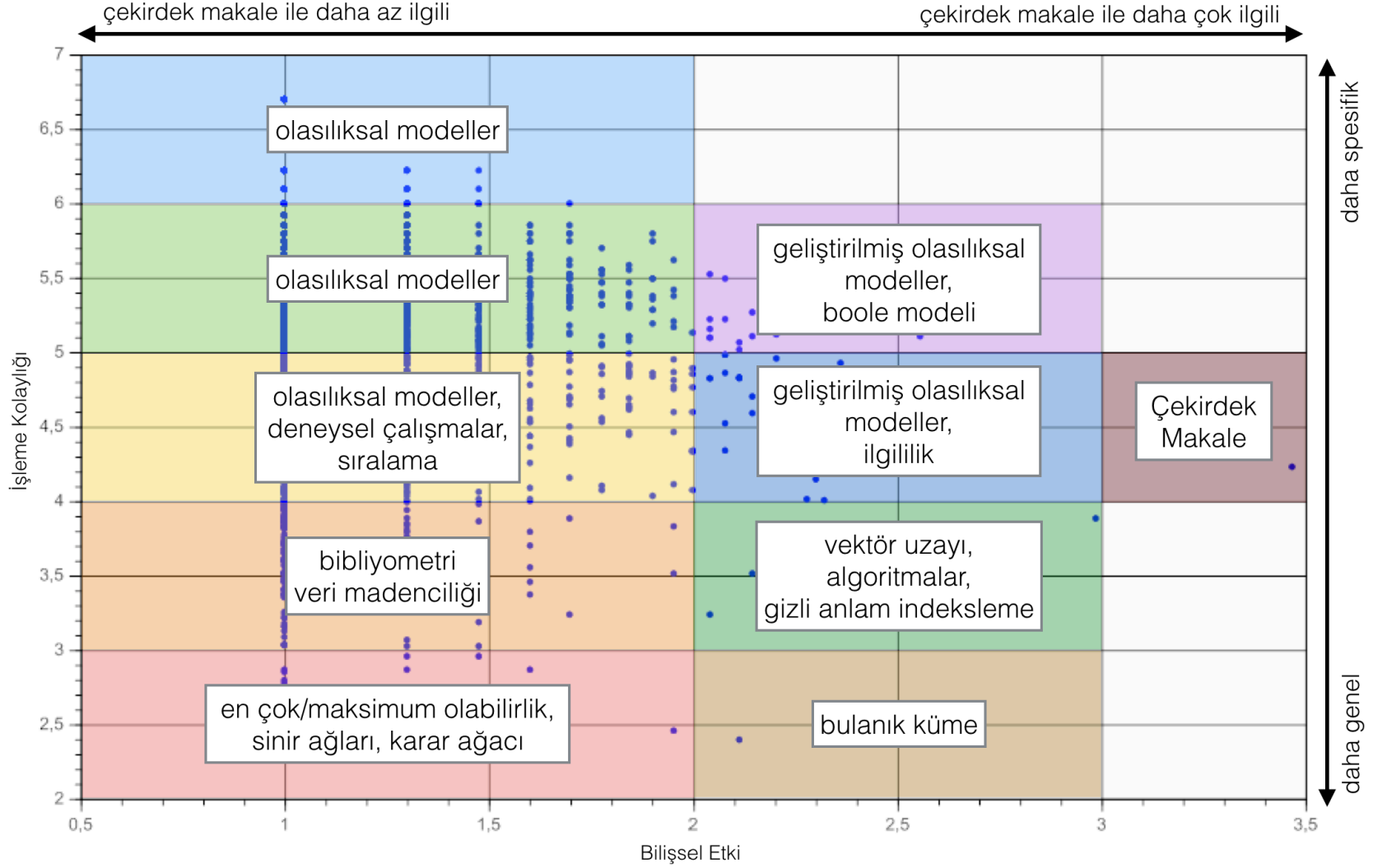
Tür olarak incelendiğinde ise A sektöründe atıf sayıları göreceli olarak daha düşük olan makaleler ve tezler görülmektedir. B sektörü orta derecede atıf

sayılarına sahip dergi makalelerini, C sektörü ise çok atıf almış makaleleri, araştırma ve değerlendirmeleri ve kitapları içermektedir

İşleme kolaylığı düşük olan kitaplar genellikle C sektöründe yer almaktadır. Bu çalışmalar daha uzun oldukları için makaleler ile kıyaslandığında hem okumak için daha fazla emek gerektirmekte (White 2007a, s. 557) hem de makalelere göre daha genel konuları ele almaktadır. Bu sebeplerden dolayı C sektöründeki çalışmaların çekirdek makale ile ilişkilendirilmeleri daha zordur. Bu bölgedeki çalışmalar genellikle dolaylı olarak çekirdek makalenin işlediği konuya değinen çalışmalardır.

Şekil 9 Maron ve Kuhns'un (1960) çekirdek makalesi için çalışmaların konusal olarak dağılımını göstermektedir. Eğer bir çalışma çekirdek makaleden daha öte bağlamlarda kullanılmışsa çekirdek makale ile ilişkisinin göreceli olarak izlenebilmesi zordur (*idf* yüksek ama *tf* düşük). C sektöründe spesifik konulara rastlamak zordur. Çalışmalar çoğunlukla genel konularla ilgilidir ve tür olarak çoğunlukla monografiler bulunmaktadır. Bu sektördeki çalışmaların B sektöründekilere göre toplam atıf sayıları daha yüksektir. Geniş kapsamları sebebiyle C grubundaki çalışmaların çekirdek çalışma ile ilgililiği tamamen açık değildir ve bu çalışmaların çekirdek makale ile gerçek ilişkilerini ortaya çıkarmak bilişsel çaba gerektirir. Genel olarak, söz konusu çalışma ne kadar çok çekirdek terimden farklı bağlamlarda kullanılırsa o çalışmayı çekirdek terim ile ilişkilendirmek o kadar zor olmaktadır. Bunun nedeni farklı bağlamlarda kullanıldıkça *tf* değerinin yüksek olma olasılığının azalmasıdır.





Şekil 9. Çalışmaların konusal olarak dağılımı

Şekil 9 aynı zamanda Maron ve Kuhns'un (1960) etki yaptığı araştırma alanları hakkında bir fikir vermektedir. Olasılıksal model konusunda çalışan bir araştırmacı için Maron ve Kuhns (1960) çekirdek makalesi için hazırlanan pennant diyagramı araştırmacıya yol gösterici niteliktedir. Pennant diyagramı hem literatür izleme açısından hem de söz konusu modelin doğrudan ya da dolaylı olarak hangi modelleri etkilediği ya da hangi modellerden etkilendiği hakkında bilgi vermektedir.

#### **4.2.2. Maron ve Kuhns'un (1960) Literatüre Etkilerinin Pennant Diyagramları ile İzlenmesi**

Bilgi erişim modelleri ile ilgili çalışmalar, çekirdek makaledeki kavramların farklı şekilde yorumlandığı çalışmalar ve ampirik verilere dayanan çalışmalar pennant diyagramında hiyerarşik olarak izlenebilmektedir.

Maron ve Kuhns'un (1960) olasılıksal dizinleme üzerine önerisi bilgi erişim literatüründe uzun yıllar tartışılmış ve birçok kişi tarafından modelin eksikleri giderilmeye çalışılmıştır. Çalışmanın literatüre etkisi kabaca iki yönlü olarak devam etmiştir. Bir taraftan çekirdek makalede ortaya atılan Model 1'in iyileştirilmesine yönelik çalışmalar yapılırken diğer taraftan da Model 1'i farklı yöntemler ve modeller ile birleştiren çalışmalar yapılmıştır. Bunun dışında çekirdek makaledeki kavramların farklı şekillerde yorumlandığı çalışmalar, ampirik verilere dayanan çalışmalar ve çekirdek makaleden önce yayımlanan çalışmalar diyagramın farklı bölümlerinde bir arada bulunmaktadır.

*"Model 1'in iyileştirilmesine yönelik çalışmalar"* incelenecek olursa: Çekirdek makalede ortaya atılan Model 1, olasılıksal parametreleri kestirmedeki problemler ve teknolojik koşullarının yetersizliği sebebi ile ilk yıllarda uygulanamamıştır. Fuhr (1986, 1989) bu parametreleri elci dizinleme ile belirleyerek bu sorunun üstesinden gelmiştir. Model 1 üzerinde geliştirmeler yapan Fuhr'un çalışmaları spesifik olarak çekirdek makale ile bire bir ilgili olduğu için A sektöründe görülmektedir. Cooper ve Maron'ın (1978) Model 1'i

formüle ederek fayda teorisine göre yorumladıkları çalışmaları da A sektöründe ve çekirdek makaleye nispeten daha yakın konumda yer almıştır.

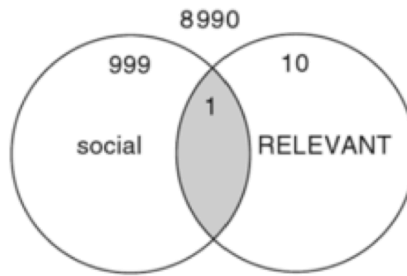
Thompson (1988) potansiyel kullanımı kestirmek zor olsa da iyi eğitilmiş bir dizincinin bir belgenin yakın gelecekte kullanımıyla ilgili mantıklı bir kestirimde bulunabileceğini belirtmektedir. Bu belgeye erişen araştırmacılar tarafından sağlanan ilgililik yargılarının zamanla dizinci tarafından atanan ilk olasılıkları etkileyeceğini de vurgulamaktadır. Model 1 ile yakından ilgili olduğu için bu çalışma çekirdek makaleye daha yakın bir noktada bulunmaktadır. Thompson'ın (1990a, 1990b) çekirdek makaledeki olasılık tahmini ile doğrudan ilgili çalışmaları da A sektöründe fakat nispeten daha uzak bir noktada bulunmaktadır. Kısaca Model 1 ile ilgili çalışmaların A sektöründe toplandığı söylenebilir.

Maron ve Kuhns'un (1960) *olasılıksal erişim modeli bilgi erişim literatüründeki hangi modelleri etkilemiş ya da hangi modellerden etkilenmiştir?* (birinci araştırma sorusu) sorusu ile ilgili olarak ise: Olasılıksal model ile yakından ilgili olan bulanık mantık (fuzzy logic) kuramı Zadeh (1965) tarafından ortaya atılmış ve su arıtma sistemlerinden borsaya kadar çok farklı alanlarda kullanılmıştır. Bulanık kümelerdeki en önemli özellik kesin olmayan sınırlardır (sadece bir kümeye ait olmama durumu). Bulanık mantık işte bu kesin olmayan tanımların matematiksel ifadesi olarak tanımlanabilir. Kuram çoklu değerlilik içermektedir. Yani geleneksel modeldeki ikili önerme yerine üç veya daha fazla sayıda önerme oluşturulmaktadır. Olasılık sıralama ilkesinde de belgeleri dizineleyen kişi belgeye ilgili veya ilgili değil şeklinde ikili (binary) ilgililik olasılığı atamak yerine, belgenin bilgi ihtiyacını karşılama olasılığını göz önünde bulundurarak 0 (ilgisiz) ile 1 (ilgili) arasında değişen bir değer atamaktadır. Bu açıdan bulanık mantık ile olasılık sıralama ilkesi yakından ilgilidir. Zadeh'nin (1965) bulanık kümeler ile ilgili yazısı C sektöründe, x ekseninin en alt bölümünde yer almaktadır. Bunun nedeni toplam atıf sayısının çok yüksek (20.046) olmasıdır.

Maron ve Kuhns'un (1960) geliştirdiği Model 1'in ardından Robertson ve Sparck Jones (1976) Model 2'yi geliştirmişlerdir. Model 1'de dizinlemeyi yapan kişinin amacı, belirli bir sorgu terimini giren kişinin belgeyi ilgili bulma olasılığını tahmin

etmektedir. Bu olasılık belgelerin hangi sorgu terimi ile dizinleneceğine karar verme durumunda önemlidir. Model 2’de ise acaba hangi sorgu cümlelerini giren kişiler belli bir terimle temsil edilen belgeleri görmekten hoşnut kalırlar sorusunun cevabı tahmin edilmeye çalışılmaktadır. Bunu yaparken de dizinci olaya Model 1’deki gibi dizinleme açısından değil de sorgu cümlesi açısından bakmaktadır. Çalışma bu açıdan çekirdek makale ile bire bir ilişkilidir ve bir atfı klasiği olduğu için C sektöründe yer almaktadır. Diğer yandan Robertson ve Sparck Jones’un çalışması (1976) diyagramda çekirdek makaleye en yakın konumda bulunan çalışmadır.

Robertson ve Sparck Jones (1976) *benzerlik ilkesine göre erişim anomalisini* fark ederek Model 2’yi geliştirmişlerdir. Göker ve Davies (2009) bu anomaliyi ayrıntılı olarak tartışmış ve aşağıdaki örneği sunmuşlardır.



Şekil 10. Sorgu terimi *social* ile Venn diyagramı (Göker ve Davies, 2009, s. 9)

Örnekte toplam 10.000 belgeden oluşan bir derlem bulunmaktadır. Bu 10.000 belgenin 1000 tanesi “social” terimi ile dizinlenmiştir. Derlemde 11 tane ilgili belge bulunmaktadır ve bu 11 belgenin sadece bir tanesi “social” terimi ile dizinlenmiştir. “Social” terimi hem sorguda hem de belgede geçiyorsa kullanıcı belgeye erişecektir. Fakat öznel ilgililik söz konusu olduğu için her “social” ile dizinlenen belgeyi kullanıcı ilgili bulmayabilir ya da “social” olarak etiketlenmeyen belgeleri ilgili bulabilir. Bu noktada “social” terimiyle dizinlenen belgelerden rastgele birine erişirse mi o belgenin ilgili olma olasılığı daha yüksek olur yoksa dizinlenmeyen belgelerden rastgele birine erişirse mi daha yüksek olur? sorusu gündeme gelmektedir.

Venn şemasındaki değerleri kullanarak soru cevaplanacak olursa:

“Social” terimi ile dizinlenen belgelerden seçilirse ilgililik oranı,  $1/1000=0,0010$  (binde bir)

“Social” terimi ile dizinlenmeyen belgelerden seçildiğinde ise ilgililik oranı  $10/9000=0,0011$  (dokuzbinde 10) olacaktır.

Dolayısıyla kullanıcı “social” terimini yazmadan rastgele bir belgeye erişse ilgili bir belgeye erişme olasılığı “social” terimi yazdığından daha yüksektir (küçük bir oran ama yüksek). Bu anomaliden yola çıkılarak Model 2 geliştirilmiştir. Her ne kadar Model 2 Robertson ve Sparck Jones (1976) tarafından geliştirilse de bu anomali ile ilgili bulgular Miller’ın (1971) çalışmasında yer almaktadır. Bu durum Model 2’nin temellerinin aslında Miller’ın (1971) çalışması ile atıldığı şeklinde yorumlanabilir. Model 2’nin ilk kez tanıtıldığı çalışmada atıf olmasa da bu ilişki pennant diyagramları ile gözlenebilmektedir. Bu bulgu araştırmanın temel hipotezi “Geleneksel atıf analizi ile gözlenemeyen disiplinlerarası ilişkiler pennant diyagramları yöntemi ile ortaya çıkarılabilir” hipotezini destekler niteliktedir. Model 2’nin oluşmasında etkili olan Miller’ın (1971) çalışması A sektöründe yer almıştır (Van Rijsbergen, 2005).

Robertson, Maron ve Cooper (1982) tarafından geliştirilen Model 3, yani birleştirilmiş olasılıksal erişim modeli (unified model) ise Model 1 ve Model 2’nin birleşimidir. Söz konusu modelde bir tarafta dizinci tarafından belgeye izin terimleri olasılıksal olarak atanırken, aynı anda kullanıcıdan sağlanan bilgiler kullanılarak terimler için olasılıksal kestirimler yapılmaktadır. Model 3’ün tanıtıldığı ve çekirdek makale ile doğrudan ilgili olan Robertson, Maron ve Cooper (1982) ve Model 3’ün geliştirilmesi için yapılan diğer çalışmaların tümü (Bodoff, 1999; Bodoff ve Robertson, 2004; Maron, 1984; Maron, Curry ve Thompson, 1986; Robertson, 2003; Thompson, 1990a, 1990b) A sektöründe yer almaktadır. Model 1’in Model 3 üzerindeki dolaylı etkisinden dolayı bu çalışmaların çekirdek makale ile spesifik olarak ilgili olan çalışmaların yer aldığı A sektöründe bulunması şaşırtıcı değildir. Model 3 ile ilgili olarak atıf klasiği

niteliğinde bir çalışma olduğu için sadece Robertson, Maron ve Cooper (1983) C sektöründe yer almaktadır.

Belgelerin ağırlıklandırılmış vektörler olarak temsil edilmesi kavramsal olarak ilk kez Maron ve Kuhns'un (1960) makalesinde kullanılmıştır. Bu açıdan da bu çekirdek makale literatürdeki ağırlıklandırma ile ilgili birçok çalışmaya kaynaklık etmiştir. Sparck Jones (1973) terimleri ağırlıklandırmak için *ters belge sıklığı (idf)* yöntemini geliştirmiştir. Salton ve Yang (1973) ise bu yöntemi olasılıksal erişim için geliştirdikleri *vektör uzayı modelinde* kullanmışlardır. Kısmi çakışmaya dayanan vektör uzayı modeli de sorgu terimlerinin ve belgelerdeki dizin terimlerinin ağırlıklandırılması ile ilgili olup (Turtle ve Croft, 1991) B sektöründe yer almaktadır. Vektör uzayı modelini konu eden diğer çalışmalar (Salton ve McGill, 1983; Salton, Wong ve Yang, 1975) ise C sektöründe bulunmaktadır.

World Wide Web'in ve web arama motorlarının yaygınlaşması ile grafik yapıları algoritmalar geliştirilmeye başlanmıştır. Bu kapsamdaki çalışmalar B sektöründe yer almaktadır. Kleinberg'in (1998) HITS ve Brin ve Page'in (1998) PageRank algoritmalarını tanıttıkları makaleler B sektöründe bir araya toplanmıştır. 2000'lere doğru literatürdeki web arama algoritmalarının gelişimi göze çarpmaktadır. Bu algoritmalarından olasılıksal mantığa dayalı model (Van Rijsbergen, Harper ve Porter, 1981) ile dil modeli (Croft ve Lafferty, 2003) A sektöründe yer almaktadır.

*“Çekirdek makalede kullanılan kavramların farklı şekillerde yorumlandığı çalışmalar”* incelendiğinde: Maron ve Kuhns'un (1960) makalesinde kullanılan kavramların farklı şekilde yorumlandığı çalışmalar A sektöründe toplanmıştır. Maron ve Kuhns'un (1960) olasılık sıralama ilkesi (probability ranking principle), iki değerli (binary) dizinlemenin yerini alan ve koleksiyonlardaki belgelerin ilgililiği hakkında tahminlerde bulunmak amacıyla istatistikleri ve kullanıcı sorgularını kullanan bir modeldir. Dolayısıyla ilgililik kavramının tartışıldığı ve farklı şekilde yorumlandığı çalışmalar (Maron, 1984; Saracevic, 1975) diyagramda diğer çalışmalara göre nispeten çekirdek makaleye daha yakın konumda yer almaktadır. Cooper ve Maron'ın (1978) öznel yararlılık açısından

İlgililiği fayda teorisine (utility theory) göre yorumladıkları çalışması ve Maron'ın (1977) hakkındalık (aboutness) kavramını ilgililiğe katkıda bulunan bir faktör olarak tartıştığı makalesi de çekirdek makaleye yakın bir konumda bulunmakta ve A sektöründe yer almaktadır. A sektöründe yer aldığı halde çekirdek makaleden nispeten daha uzak konumda bulunan çalışmalar ise Fairthorne'un (1961) bilgi erişim ile ilgili genel kapsamdaki makalesi ve Van Rijsbergen'in (2005) olasılıksal bilgiye erişimin kökleri hakkındaki makalesidir.

Her ne kadar Maron ve Kuhns (1960) ilgililik geribildirim (relevance feedback) terimini kullanmamış olsalar da, bu kavramı ilgili ve erişilen belgelerin kümesini genişletme tekniklerini tartışırken üstü örtülü olarak açıklamışlardır. Rocchio'nun (1966) ilgililik geribildirim ile ilgili çalışması da bu nedenle A sektöründe yer almaktadır. Çekirdek makale ile ilgili olan bir diğer kavram ise olasılıksal sıralama prensibidir. Cooper (1978) olasılıksal sıralama prensibinden resmi olarak ilk kez bahseden kişidir, ancak prensibin kullanımı Maron ve Kuhns'un (1960) makalesine kadar geri gitmektedir. Olasılıksal sıralama prensibinin özünde çeşitli kaynaklardan toplanan bilgilerin arama sonuçlarını geliştirmek için kullanılması bulunmaktadır. Bu iki çalışma da A sektöründe diğer ampirik verilere dayanan kalite testleri (benchmark collections) çalışmaları ile birlikte yer almaktadır.

Maron ve Kuhns (1960) aynı zamanda belgeler arasındaki benzerliklerin nasıl belirlenebileceğini de göstermişlerdir. İndeks terimleri ya da belgeler arasında bu benzerliklere bakmak o dönem için alışılmışın dışındadır. O günden bu yana pek çok araştırmacı dizin terimleri ile belgeler ya da belgeler ile belgeler arasında benzerlik ölçüleri geliştirmiştir. Vektör uzayı modeli ile ilgili olan bu tür çalışmalardan Salton, Wong ve Yang (1975) C sektöründe, Turtle ve Croft (1991) ise B sektöründe yer almaktadır.

*"Ampirik verilere dayanan çalışmalar"* ile ilgili olarak: Dizinlemeye farklı bir bakış açısı getiren Maron ve Kuhns'un (1960) makalesi, o zamanki teknoloji koşulları modelin uygulanmasına kolayca imkân vermediği için sadece kuramsal bir yaklaşım içermektedir. Bu yüzden model ilerleyen yıllarda düşünce deneylerinde kavramsallaştırılmaya çalışılmıştır. Daha önce de değinildiği gibi,

Maron ve Kuhns'un (1960) olasılık sıralama ilkesinin özünde çeşitli kaynaklardan toplanan bilgilerin arama sonuçlarını geliştirmek için kullanılması bulunmaktadır. Bu durum ampirik verilere dayanan çalışmalara öncülük etmiştir. Good'un (1958) olasılıksal bilgi erişimi tartıştığı çalışması ve konuyla ilgili ampirik verilere dayanan bir çalışma olan Luhn (1961) A sektöründe yer almaktadır. Olasılıksal sıralama prensibinin kanıtlandığı diğer çalışmalar (Robertson, 1977; Croft, 2000; Salton ve Lesk, 1968; Saracevic ve Kantor, 1988a, 1988b; Saracevic, Kantor, Chamis ve Trivison, 1988) ile Turtle ve Croft'ın (1991) sıralı erişim modeline geçiş aşaması ile ilgili çalışması B sektöründe bulunmaktadır.

Bir diğer ampirik verilere dayanan çalışma ise Blair ve Maron'ın (1985) Bechtel firması tarafından San Francisco Körfez Bölgesinde yapılan metro inşaatı sırasında meydana gelen bir kaza üzerine açılan davada kullanılan bir derlem üzerinde gerçekleştirdikleri performans değerlendirmesiyle ilgilidir. Bu ampirik verilere dayanan çalışma ise B sektöründe yer almaktadır.

*"Model 1'den Önceki Çalışmalar"* incelendiğinde: Maron ve Kuhns (1960) çekirdek makalesi için hazırlanan pennant diyagramında çekirdek makaleden daha önceki yıllarda hazırlanmış çalışmalar da yer almaktadır. Söz konusu çalışmalar türlerine ve konularına göre farklılık gösterdiği için tüm sektörler yayılmıştır. Örneğin, bilgisayarlar için düşünülen ilk kullanım alanlarından biri olan belge erişimi ile ilgili literatürdeki ilk çalışma olan Bagley (1951) ve Good'un (1958) olasılıksal bilgiye erişimi kısaca tartıştığı IBM teknik raporu A sektöründe yer almaktadır. Çekirdek makalenin olasılıksal bilgi erişim ile ilgili olduğu düşünülürse bu iki çalışmanın çekirdek makale ile en ilgili çalışmaların yer aldığı çalışmaların bulunduğu sektörde yer alması şaşırtıcı değildir.

İnsan beyinde dizinleme ve erişim ilişkili dizinleme (associative indexing) yoluyla yapılmaktadır (Bush, 1945, Tonta, 2002). Vannevar Bush'un bu yapıdan yola çıkarak *memex* adını verdiği aleti bir kimsenin tüm kitaplarını, resimlerini, yazışmalarını vb. mikrofilm olarak depoladıktan sonra aralarında ilişki kurulabileceği ve gerektiğinde kolaylıkla sorgulanabileceği bir aygıt olarak tanımlamıştır (Tonta, 2001). *Hipermetin* (Hypertext) terimini kullanmadan



“hipermetin”in tanımlandığı bu makale B sektöründe yer almaktadır. Bush’un ilişkili dizinleme (associative indexing) yöntemi ile ilgili bu çalışmada çekirdek makalede konusu geçen bilimsel makaleleri bulma problemi tanımlanmıştır. Bu yüzden ilgililikleri açık olarak belli olmamasına rağmen diyagramdaki konumu incelendiğinde çekirdek makaleye daha yakın bir konumda olduğu görülmektedir.

Shannon ve Weaver’ın (1949) Matematiksel İletişim Kuramını tanımladıkları makalesi konuyla ilgili çok sayıda çalışmaya kaynaklık etmiştir. Bu çalışma aynı zamanda çekirdek makalenin kaynakçasındaki iki çalışmadan biridir. Çalışma atıf klasiği niteliğinde olmasından dolayı C sektöründe yer almaktadır.

Deerwester, Dumais, Furnas, Landauer, ve Harshman’ın (1990) gizli anlam dizinleme ile ilgili çalışmaları da çok fazla atıf alan çalışmalardan biri olarak C sektöründeki yerini almıştır. Daha önce de belirtildiği gibi C sektöründe genellikle çok yüksek atıf alan ve eski tarihlerde yayımlanan çalışmalar yer almaktadır. Çalışmaların yazarları alanda büyük katkılar yapmış ünlü kişilerdir. Bulanık kümeler teorisinin sahibi Zadeh (1965), bibliyometri alanında kendi adı ile anılan yasası bulunan (Price Law) Derek de Solla Price (1965), atıf dizinlerini ortaya atan Garfield (1972), modeller ile ilgili olarak Akaike bilgi ölçütünü (Akaike information criterion-AIC) geliştiren Akaike (1974), 2012 Nobel Ekonomi ödülü sahibi Kahnemann’la birlikte "Beklenti Teorisi"ni geliştiren Tversky’nin çalışmasının (Tversky ve Kahneman, 1974) C sektöründe yer alması şaşırtıcı değildir.

#### 4.2.3. Pennant Diyagramları ve İlgililik Puanları

Pennant diyagramları yöntemi çekirdek terim (seed) ile birlikte geçen terimlerin ya da birlikte atıf yapılan yazarların ilgililik puanlamasını hesaplamak için de kullanılmaktadır (White, 2007a, 2007b, 2010). Bu değer normalleştirilmiş  $tf$  ve  $idf$  değerlerinin çarpımından elde edilir ( $1+\log(tf)*\log(N/df)$ ). Yüksek ilgililik puanı  $tf$  ve  $idf$  değerinin birbirine yakın olması anlamına gelmekte ve ilgili çalışmanın çekirdek makaleye yakın olarak konumlandırılmasını sağlamaktadır.

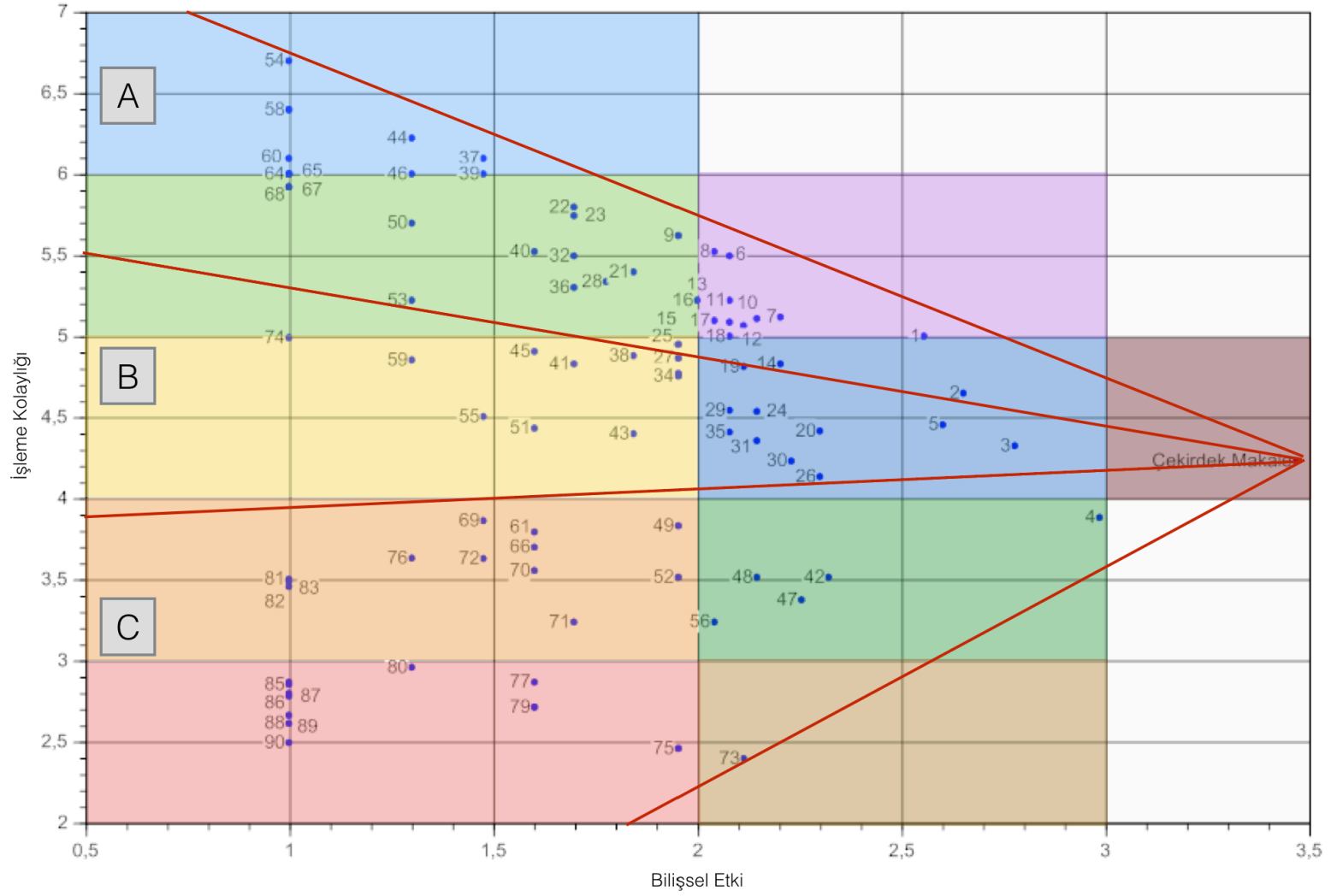
Pennant diyagramı yöntemi ile hesaplanan ilgililik sıralamasında en yukarıda olan çalışmalar (ilgililik puanı en yüksek olanlar) genellikle A ve B sektörlerinde çekirdek makaleye en yakın bölgede toplanmıştır (Şekil 11). Diğer bir deyişle bu bölgedeki kaynaklar hem bilişsel etki hem de işleme kolaylığı skalalarında en yüksek puanı almış olan çalışmalardır.

Pennant diyagramları yönteminde ilgililik sıralaması hesaplanırken ortak atıf değerinden elde edilen bilişsel etki ölçüsü de hesaba katılmaktadır. Dolayısıyla işleme kolaylığı skalasında daha aşağıda olan öğelerin ilgililik sıralamasında ilk sıralarda olabilmesi için bilişsel etki skalasında da ilk sıralarda olması gerekir.

Formül 3'teki *idf* faktörü *çekirdek* ile daha az ilgili öğeleri aşağı itmekte ve daha ilgili olanları ise yukarı taşımaktadır. *Idf* etkisinin ne kadar başarılı olduğunu görmek için başlığında "olasılıksal" terimi geçen çalışmalar için hazırlanmış olan etkileşimli grafik<sup>17</sup> incelenebilir. Söz konusu çalışmaların hemen hepsi A sektöründe toplanmıştır (%81). Yine de *tf\*idf* değerine göre ağırlıklandırma her alanda atıf verme biçimleri farklı olduğu için farklı alanlardan çalışmalar için farklı sonuçlar verebilir.

---

<sup>17</sup> Bkz. <https://goo.gl/h4RmpD>

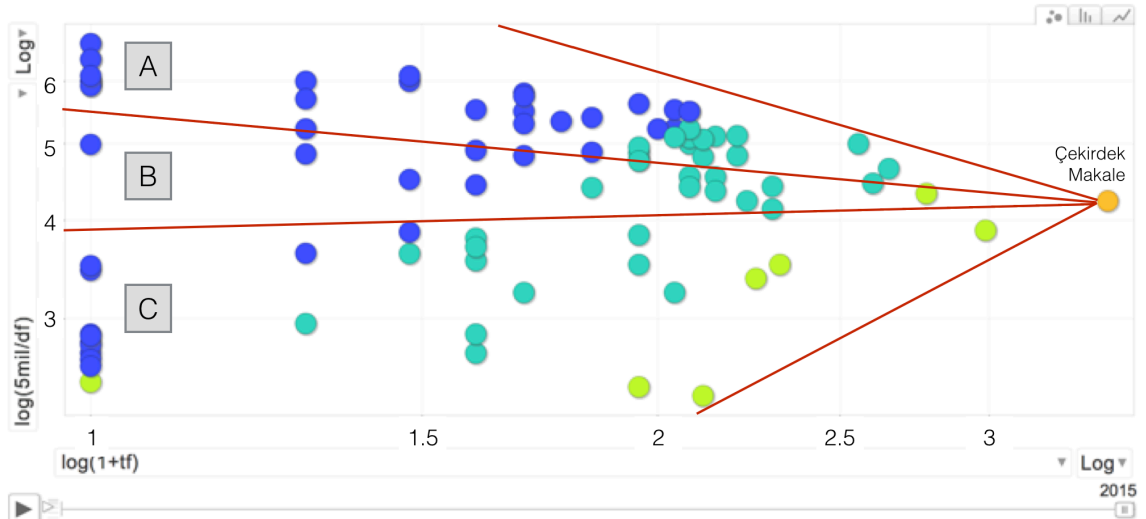


Şekil 11. İlgililik sıralamalarının pennant diyagramı gösterimi

### 4.3. MARON VE KUHNS'UN (1960) ÇEKİRDEK MAKALESİ İÇİN ETKİLEŞİMLİ PENNANT DİYAGRAMLARI

Çalışma kapsamında hazırlanan etkileşimli pennant diyagramları (<https://goo.gl/Wjr1YD>) statik pennant diyagramlarından farklı olarak atıfların yıllara göre dağılımı ile ilgili bilgi sağlamaktadır. Etkileşimli diyagramda çekirdek makalenin yayınlandığı 1960 yılından 2015 yılına kadar tüm senelerdeki değişim izlenebilmektedir. Diyagramın sağ tarafındaki panellerde görünüm ile ilgili değişiklikler yapılabilmektedir. Diyagramlardaki noktaların renkleri ve büyüklükleri ile x ve y eksenlerinde yer alması istenen değerler isteğe göre değiştirilebilmektedir. Belirli çalışmalar seçilerek sadece o çalışmaların etkileşimleri izlenebilmektedir. Diyagramın üst bölümündeki sekmeler sayesinde de saçılım grafiği sütun ya da çizgi grafiğe dönüştürülebilmektedir.

Etkileşimli diyagram çekirdek makale ile en az bir kere birlikte anılmış olan 4176 kaynak üzerinden kolayda örneklem yöntemi ile seçilen 90 kaynak (çekirdek makale hariç) için hazırlanmıştır.<sup>18</sup> Etkileşimli grafiğin sonuç görüntüsü Şekil 12'teki gibidir.



Şekil 12. Maron ve Kuhns (1960) etkileşimli pennant görüntüsü

<sup>18</sup> Doksan kaynaktan oluşan etkileşimli pennant diyagramı için bkz. <https://goo.gl/R2VXVm> Diyagramları oluşturmak için kullanılan veri setleri de aynı dosyada bulunmaktadır.

Grafikte herhangi bir noktanın üzerine tıklanarak söz konusu çalışmanın yıldan yıla çekirdek makaleye göre olan konum değişiklikleri gözlenebilir. Diyagramlar çalıştırıldığında ilk yıllarda nokta sayısı daha azdır. Örneğin çekirdek makale yayımlandıktan dört yıl sonra (1964) diyagramda sadece iki nokta bulunmaktadır çekirdek çalışma hariç). İlerleyen yıllarda çekirdek makale ile birlikte anılan makaleler diyagrama dahil olmaya başlamaktadır. Grafik çalıştırıldığında bazı noktaların belli yıllarda aşağı doğru inerken çekirdeğe doğru yaklaşmamaları o çalışmanın atıf almaya devam ettiği fakat çekirdek makale ile ortak atıf almadığı şeklinde yorumlanabilir. Çekirdeğe doğru ilerleyen çalışmalar ise hem atıf almaya devam eden hem de çekirdek makale ile birlikte anılan çalışmalardır. Başka bir deyişle, bu çalışmaların çekirdek çalışmayla benzer konularda olma olasılığı yüksektir. İsteğe göre noktaların büyüklükleri ilgililik ağırlığına göre düzenlenebilir.

Grafik çalıştırıldığında noktaların çekirdek makaleye yaklaşma hızları birlikte anılma oranları hakkında bilgi vermektedir. Belli yıl aralıklarında hızla çekirdek makaleye yaklaşan noktalar o dönemde birlikte atıf sayısının arttığı anlamına gelmektedir. Öte yandan noktaların aşağı doğru hızlandığı fakat çekirdek makaleye doğru yavaş hareket ettiği durumlar ise ilgili çalışmaların toplam atıf sayısının arttığı fakat çekirdek makale ile çok fazla birlikte anılmadığı anlamına gelmektedir. Bu bağlamda genel olarak Model 1'in 1980'lerin ilk yarısında sadece olasılıksal modeller ile anıldığı, ikinci yarısında ise olasılıksal sıralama ve ilgililik konuları ile ilgili çalışmaları etkilediği gözlenmiştir. Gizli anlam dizinleme ile ilgili çalışma ise yayımlandıktan 14 yıl sonra çekirdek makale ile ortak atıf almaya başlamış ve birlikte anılma sayıları son altı yılda giderek artmıştır. Bu bulgular *"Etkileşimli pennant diyagramlarında olasılıksal erişim modelinin hangi yıllar aralığında hangi modelleri etkilediği gözlenebilir mi?"* sorusunun cevabı niteliğindedir (üçüncü araştırma sorusu)

#### **4.4. İLGİLİLİK SIRALAMALARI**

Çalışma kapsamında Maron ve Kuhns'un (1960) çekirdek makalesi için WoS ilgili kayıtlar yöntemi ve pennant diyagramları yöntemi ile olmak üzere iki ayrı ilgililik sıralaması oluşturulmuştur.

Çekirdek makale için kayıt sayısı pennant yöntemi ile 4176, WoS ilgili kayıtlar yöntemi ile 9803 olarak bulunmuştur. Değerlendirmenin daha sağlıklı yapılabilmesi için ilgililik sıralamasındaki ilk 50 kayıt ayrıntılı olarak incelenmiştir (makalelerin başlıkları için bkz. Ek 3).

İki yöntemle ilgililiğe göre sıralanan ilk 50 kaydı ele aldığımızda, iki listenin birbirini kapsama oranı %0'dır. Yani pennant diyagramı yöntemi ile oluşturulan sıralamaya göre en ilgili 50 başlığın hiçbirisi WoS sıralamasına göre ilk 50 başlığa dahil değildir. Bu bulgu çalışmadaki dördüncü araştırma sorusu olan "Pennant diyagramları yöntemi ile elde edilen *ilgililik sıralaması* ile WoS veri tabanındaki ilgili kayıtlar (related records) özelliği ile elde edilen ilgililik sıralaması ne kadar örtüşmektedir?" sorusunun cevabıdır.

İlgililik sıralamalarındaki ilk 50'şer kaydın karşılaştırıldığı Ek 3 incelendiğinde WoS ilgililik sıralamasında 12. kayıttan sonraki çalışmaların hepsinin 2015 yılında yayımlandığı görülmektedir. Bunun nedeni WoS'un kaynakçadaki eşleşen kayıt sayısı aynı olan çalışmalar için güncel olanları başta göstermesidir.

WoS'un ilgili kayıtlar özelliği ile oluşturulan ilgililik sıralaması sadece kaynakça benzerlikleri dikkate alınarak yapılmaktadır. Dolayısıyla bu yöntem kaynakçası zengin olmayan kaynaklar için sağlıklı olmayan, yanıltıcı sıralamalar üretmektedir. Maron ve Kuhns'un (1960) makalesinin kaynakçasında sadece iki çalışma bulunduğu düşünüldüğünde kaynakça benzerliğine göre yapılan ilgililik sıralamasının sağlıklı sonuçlar vermeyeceği açıktır.

Tablo 5, WoS ilgili kayıtlar özelliğine göre oluşturulan listedeki kaynakların araştırma alanlarına göre dağılımı ile ilgili bilgi vermektedir. Söz konusu araştırma alanları WoS tarafından belirlenmiştir (SU=Research Area). Tablo incelendiğinde WoS ilgili kayıtlar sıralaması ile oluşturulan listedeki kaynakların en çok bilgisayar bilimleri ile ilgili olduğu görülmektedir. İkinci sırada ise 883 kaynak ile mühendislik alanı ile ilgili kaynaklar yer almaktadır. Çok çeşitli disiplinlerden kaynakları içerdiği görülen listenin orijinal makalenin yapısına uygun olmadığı gözlenmiştir. WoS ilgili kayıtlar yöntemi ile oluşturulan listedeki

kaynak sayısı daha önce de belirtildiği gibi 9803'tür, fakat tablodaki kaynakların sayısı 9914'tür. Tablodaki sayının fazla olmasının nedeni bazı çalışmaların birden çok çalışma alanına dahil olmasıdır.

WoS ilgili kayıtlar özelliğine göre oluşturulan sıralamada benzerlik oranı en yüksek gözükten makalenin (Prevedelli, Simonini ve Ansaloni, 2001) kaynakçasında 23 kaynak bulunmakta ve bunlardan iki tanesi çekirdek makalenin de kaynakçasında yer almaktadır. Çekirdek makalenin kaynakçasında zaten iki kaynak bulunmaktadır. Dolayısıyla bu hesaplama yöntemi ile sistem çakışma oranı %100 olan bu çalışmayı en ilgili diye sıralamanın en başında göstermektedir. Oysa söz konusu makale tatlı su biyolojisi ile ilgilidir. İlgililik sıralamaları ile ilgili bu sonuçlar "Ortak atıf sayılarına göre oluşturulan ilgililik sıralamaları kaynakça benzerliğine göre oluşturulan ilgililik sıralamasından daha başarılıdır" hipotezini doğrulamaktadır.

Tablo 5. WoS ilgili kayıtlar özelliğine göre oluşturulan listedeki kaynakların araştırma alanlarına göre dağılımı

Sıra No	Araştırma Alanı	Kaç Kaynakta Geçiyor
1	Bilgisayar Bilimi	1336
2	Mühendislik	883
3	Deniz ve Tatlı Su Biyolojisi	744
4	Psikoloji	732
5	Fizik	719
6	İşletme Ekonomisi	488
7	Bitkibilim	439
8	Matematik	432
9	Ziraat	427
10	Zooloji	405
11	Kimya	336
12	Nöroloji	333
13	Kütüphanecilik ve Bilgibilim	325
14	Bilim Teknoloji Diğer Konular	255
15	Biyçeşitliliğin Korunması	254
16	Yaşam Bilimleri Diğer Konular	244
17	Biyokimya Moleküler biyoloji	228
18	Okyanus Coğrafyası	226
19	Ormancılık	212
20	İletişim	208
21	Mikrobiyoloji	187
22	Jeoloji	183
23	Dil Bilimi	164
25	Hesaplamalı Biyoloji	154

Çekirdek makalede atıfta bulunulan iki kaynak matematiksel iletişim kuramı (Shannon ve Weaver, 1949) ve önermeler arasındaki olası ilişki katsayıları (Yule, 1912) ile ilgili çalışmalardır. WoS ilgili kayıtlar özelliğinde ilgililik sıralaması daha çok çekirdek makalenin kaynakçasına göre şekillenmektedir. Pennant sıralamasında ise çekirdek makalenin yazarının kaynakçası dışında diğer araştırmacıların atıfları da hesaplamaaya dahil edilmektedir. Dolayısıyla ilgililik sıralamasında çalışmaların literatürdeki etkileri de hesaba katılmaktadır. Pennant yöntemi ile hesaplanan sıralamanın çekirdek makalenin yapısına daha uygun olduğu gözlenmiştir (bkz. Ek 3). Bu yöntem ile elde edilen sıralamada bilgi erişim modelleri ağırlıktadır. Sıralama dikkatle incelendiğinde ise olasılıksal model ile ilgili çalışmaların sıralamada daha üst sıralarda yer aldığı göze çarpmaktadır. Sıralamadaki ilk beş çalışma olasılıksal model ve ilgililik ile ilgili konularda yazılmıştır (ilgililik sıralarına göre: Cooper ve Maron, 1978; Robertson, Maron ve Cooper, 1982; Robertson, 1977; Robertson ve Sparck Jones, 1976; Croft ve Harper, 1979). Bu bulgu araştırmanın ikinci sorusunun cevabı niteliğindedir.

Maron ve Kuhns'un (1960) bilgi erişim literatürüne katkısı başka araştırmacıların da ilgisini çekmiştir. Thompson (2007) *Information Processing and Management* dergisinin özel sayısında yayımladığı çalışmasında söz konusu çalışmanın bilgi erişim literatürünü nasıl etkilediğini tartışmıştır. Thompson'ın referans listesinde 51 çalışma bulunmaktadır. Bu çalışmalardan 41 tanesi bu çalışma kapsamında örneklem olarak seçilen veri seti içerisine dahildir. Bu istatistik de pennant diyagramları yönteminin ne kadar başarılı sonuçlar verdiğini açıklar niteliktedir.

Pennant diyagramlarının daha başarılı sonuçlar vermesinin sebebinin alandaki diğer araştırmacıların atıflarının da hesaplamaaya dahil edilmesi olduğu düşünülmektedir. Bu sayede kendinden önceki çalışmalara referans vermemiş olsa bile yazarların hangi kaynaklardan yararlandıkları ortaya çıkmaktadır. Çalışmada adı geçen "belirli olmayan ilişkiler" diğer yazarların atıfları ile ortaya çıkmaktadır. Benzer şekilde örneğin yeni ortaya atılan bir modelin anlatıldığı bir makale başlangıç niteliğinde olduğu için kaynakçası makale içeriği ile birebir örtüşmeyebilir. Bu gibi durumlarda ileriki tarihlerde diğer araştırmacıların



yaptıkları atıflar o modelin hangi modelleri etkilediğini belirlemede yol gösterici olacaktır.

## 5. BÖLÜM: SONUÇ VE ÖNERİLER

### 5.1. SONUÇ

Etki belirleme çalışmalarının özellikle geriye dönüşlü olarak gerçekleştirilmesi önemlidir. Bilimsel alanların örüntüleri birbirinden çok farklıdır ve bu tür çalışmalar bilimsel iletişimin gerçekte nasıl işlediğini görmek açısından önemlidir. Atıf dizinlerindeki göstergelerin (h-indeks, atıf sayıları vb.) bilim insanlarının ve bilimsel çalışmaların alana katkılarını tam olarak belirleyemediği göz önüne alındığında etki belirleme çalışmalarında yeni yöntemlerin kullanılma gereksinimi ortaya çıkmıştır.

Bu çalışmada Maron ve Kuhns'un (1960) olasılıksal erişim ile ilgili çalışmasının literatüre olan etkisi pennant diyagramları yöntemi kullanılarak yorumlanmış ve önemli bulgular elde edilmiştir. Bulgular hipotezleri destekler niteliktedir. Bilgi erişim alanında olasılıksal model ile ilgili çalışan yazarlar ve modelin etkilediği ya da etkilendiği çalışmalar bir araya getirilmiştir. Kendinden önceki çalışmalara referans vermemiş olsa bile yazarların hangi kaynaklardan yararlandıkları ortaya çıkarılmıştır. Bilgi erişim literatürü Maron ve Kuhns'un olasılıksal erişim ile ilgili makalesini yayımladıkları 1960 yılından daha eskilere dayanmaktadır. Çalışmada Maron ve Kuhns'un (1960) çalışmasının daha önceki çalışmalar ile ilişkileri ortaya çıkarılmış ve bu çalışmanın hangi yapıların/modellerin oluşmasında etkili olduğu gözlenmiştir. Hazırlanan etkileşimli zaman çizelgesi ile bu ilişkilerin zamana göre değişimi gözlenerek bilgi erişim alanındaki yönelimler belirlenmiştir. Pennant diyagramları yorumlanarak bilgi erişim modellerinin gelişimi tarihsel olarak izlenebilmiş ve dolaylı bir bilgi erişim tarihine başlangıç örüntüsü ortaya konulmuştur.

Atıf dizinlerinde kaynakça benzerliğine dayalı yöntem ile oluşturulan *ilgililik sıralamaları* özellikle kaynakçası zengin olmayan çalışmalar söz konusu olduğunda iyi sonuç vermemektedir. Ortak atıf değerlerinin de hesaba katıldığı pennant diyagramları yöntemi ile daha iyi ilgililik sıralamaları oluşturulabilmektedir.

Bu çalışma literatürde pennant diyagramı yönteminin ilgililik sıralamaları oluşturulmak amacıyla kullanıldığı ilk çalışmadır. Bunun dışında literatürde pennant diyagramları yöntemi ile oluşturulan ilgililik sıralamasının atıf dizinlerinde kullanılan mevcut uygulamalarla karşılaştırıldığı bir çalışma bulunmamaktadır.

Sonuç olarak pennant diyagramı yöntemi ile elde edilen grafikler ve ilgililik sıralamaları araştırmacılar için faydalı bilgiler içermektedir. Bu yöntem sayesinde hem araştırmacılara literatür izleme konusunda kolaylık sağlanmakta hem de çekirdek makalenin yapısına daha uygun ilgililik sıralamaları oluşturulabilmektedir.

## 5.2. ÖNERİLER

Pennant diyagramları yöntemi kullanılarak üretilen ilgililik sıralamaları ve diyagramlar atıf veri tabanlarına entegre edilebilir. Sadece arayüzde değişiklik yapılarak veri tabanında mevcut bilgiler (atıf ve ortak atıf değerleri) kullanılıp kullanıcılara daha isabetli ilgililik sıralamaları sunulabilir. Buna ek olarak özellikle alanda yeni olan araştırmacılara yol gösterici nitelikte grafikler oluşturulabilir ve ilgilenilen çalışmanın doğrudan ya da dolaylı olarak hangi çalışmaları etkilediği ya da hangi çalışmalardan etkilendiği hakkında kullanıcılar bilgilendirilebilir. Bu değişiklikler atıf veri tabanlarında yapılabileceği gibi, web uygulaması aracılığı ile atıf dizinlerinden çekilen veriler ile de benzer sonuçlar elde edilebilir. Diyagramlar ya da ilgililik sıralamaları oluşturulmadan önce veri temizleme aşamasının eklenmesi açısından ayrı bir web uygulaması kullanmak daha kolay olacaktır. Burada veri temizleme aşaması örüntünün ya da sıralamanın doğru çıkması açısından en kritik süreçtir.

Herhangi bir alana ait örüntünün belirlenebilmesi için çekirdek makale seçimi önemlidir. Bu amaçla alana ait özgün makaleler ve atıf klasikleri belirlendikten sonra (örneğin Referans Yılı Spektroskopisi uygulaması ile) pennant diyagramları ya da ilgililik sıralamalarının oluşturulmasının faydalı olacağı düşünülmektedir.

Pennant diyagramı yöntemi üniversitelerin ya da kurumların hangi alanlarda etkili oldukları ve diğer kurumlarla ilişkilerinin belirlenmesi için de kullanılabilir. Potansiyel işbirliği yapılacak kurum ya da kişi seçiminde yol gösterici olabilir.

Yapılan çalışmalar belli dönemlerde tekrarlanarak sonuçlar karşılaştırılabilir ve yönelimler belirlenebilir.

### **5.3. ÇALIŞMANIN SINIRLILIKLARI VE GELECEKTE YAPILMASI ÖNERİLEN ÇALIŞMALAR**

Bu çalışma tek bir atıf klasiğine (Maron ve Kuhns (1960) dayandığı için genelleme yapmak pek doğru olmayacaktır. Fakat, yine de, olasılıksal erişim ile ilgili bir çalışmanın literatürdeki etkisinin pennant diyagramları ile yorumlandığı bu çalışmada bilgi erişim tarihi açısından kanımızca önemli bulgular elde edilmiştir. Aynı yöntem kullanılarak oluşturulan ilgililik sıralaması *Web of Science*'taki mevcut uygulamayla karşılaştırıldığında ise daha isabetli ilgililik sıralaması oluşturulduğu gözlenmiştir.

Bu çalışma uluslararası atıf dizinlerinde (*Scopus*, *Web of Science* vb.) indekslenmeyen çalışmaların da dahil edilebilmesi için Google Akademik (Google Scholar) üzerinde tekrarlanabilir.

İlgili kaynağa atıf yapan kaynaklardan WoS'ta dizinlenmemiş olanlar (dış kaynaklar) veri setine dahil edilebilir ve onların etkisi de gözlenebilir.

Çalışma aynı çekirdek makale için farklı örneklem setleri kullanılarak ya da örneklem sayısı artırılarak tekrar edilebilir. Çünkü küçük gruplarda aykırı değer (outlier) çıkma olasılığı daha fazladır ve rastlantısal bir olayı sistematik olarak sınıflamak gibi bir yanılgıya (model yanılması) düşülebilir (Kahneman, 2011, s.137).

İlgililik sıralamaları ile ilgili olarak klasiklerin (özellikle kaynakçası zengin olan çalışmalar) WoS ilgili kayıtlar yöntemi ve pennant diyagramları kullanılarak oluşturulan sıralamaları Spearman (ya da Pearson) sıra ilişki katsayısı ile karşılaştırılabilir.

Atıf sayısına dayanan pennant diyagramlarında çekirdek makaleye atıf yapan her kaynağın değeri eşittir. Oysaki, örneğin, PageRank algoritmasıyla atıflar dergilerin etki faktörlerine göre ağırlıklandırılarak pennant diyagramları oluşturulabilir.

## KAYNAKÇA

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716-723.
- Al, U. (2008). *Türkiye'nin bilimsel yayın politikası: Atıf dizinlerine dayalı bibliyometrik bir yaklaşım*. (Doktora Tezi, Hacettepe Üniversitesi, Ankara).  
Erişim adresi:  
<http://yunus.hacettepe.edu.tr/~umutal/publications/dissertation/full.pdf>
- Al, U., Sezen, U. ve Soydal, İ. (2012). *Türkiye'nin bilimsel yayınlarının sosyal ağ analizi yöntemiyle değerlendirilmesi*. (TÜBİTAK Sosyal Bilimler Araştırma Grubu - Proje No: SOBAG 110K044). Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü. Ankara. Erişim Adresi:  
<http://www.bby.hacettepe.edu.tr/bilgibelge/file/SOBAG-110K044.pdf>
- Al, U., ve Tonta, Y. (2004). Atıf analizi: Hacettepe Üniversitesi Kütüphanecilik Bölümü tezlerinde atıf yapılan kaynaklar. *Bilgi Dünyası*, 5(1),19-47. Erişim adresi: <http://bd.org.tr/index.php/bd/article/view/220/206>
- Allott, N. (2013). Relevance theory. A. Capone, F. Lo Piparo ve M. Carapezza (Yay. haz.). *Perspectives on pragmatics and philosophy* içinde (s. 57-98). Berlin/New York: Springer.
- Ateş, T. K., Özkan, S., Soysal, M. ve Alatan, A. A. (2011). Relevance feedback for semantic classification: A comparative study. *2011 IEEE 19th Signal Processing and Communications Applications Conference (SIU)*, 22-24 Haziran 2011, Ontario, Canada. Erişim adresi:  
<http://www.junjan.es.com/files/research/relevance-feedback-for-semantic-classification-a-comparative-study.pdf>
- Bagley, P. (1951). *Electronic digital machines for high-speed information searching*. (Yüksek Lisans Tezi, Massachusetts Teknoloji Enstitüsü, Cambridge). Erişim adresi:  
<http://dspace.mit.edu/handle/1721.1/12185#files-area>

- Bauer, K. ve Bakalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*, 11(9). Erişim adresi: <http://dlib.org/dlib/september05/bauer/09bauer.html>
- Beel, J., Gipp, B., Langer, S., ve Breitinger, C. (2015). Research-paper recommender systems: A literature survey. *International Journal on Digital Libraries*, 73(2), 1-34.
- Blair, D. C. ve Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Communications of the ACM*, 28(3), 289–299.
- Brin, S. ve Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7), 107-117. doi: 10.1016/s0169-7552(98)00110-x
- Bodoff, D. (1999). A re-unification of two competing models for document retrieval. *Journal of the American Society for Information Science*, 50(1), 49–64.
- Bodoff, D. ve Robertson, S. E. (2004). New unified probabilistic model. *Journal of the American Society for Information Science and Technology*, 55(6), 471–487.
- Bookstein, A. (1983). Outline of a general probabilistic retrieval model. *Journal of Documentation*, 39(2), 63-72.
- Buchanan, L. ve O Connell, A. (2006). A brief history of decision making. *Harvard Business Review*, 84(1), 32-36. Erişim adresi: <http://thegrbluebook.com/wp-content/uploads/2013/02/A-Brief-History-of-Decision-Making-by-Leigh-Buchanan-and-Andrew-OConnell.pdf>
- Bush, V. (1945). As we may think. *Atlantic Monthly*, 176(1), 101-108. Erişim adresi: <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

- Buzydlowski, J. (2002). *A comparison of self-organizing maps and pathfinder networks for the mapping of co-cited authors*. (Doktora Tezi, Drexel Üniversitesi, Philadelphia). Erişim adresi: <http://opim.wharton.upenn.edu/~sok/papers/b/janbuz-thesis.pdf>
- Carevic, Z. ve Mayr, P. (2014). Recommender systems using pennant diagrams in digital libraries. (*arXiv:1407.7276*). Erişim adresi: <http://arxiv.org/pdf/1407.7276v1.pdf>
- Carevic, Z. ve Schaer, P. (2014). On the connection between citation-based and topical relevance ranking: Results of a pretest using iSearch. *Proceedings of the first workshop on bibliometric-enhanced information retrieval* içinde (s. 37-44). Amsterdam, The Netherlands. Erişim adresi: <http://ceur-ws.org/Vol-1143/paper5.pdf>
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57, 359–377.
- Chisholm, E. ve Kolda, T. G. (1999). *New term weighting formulas for the vector space method in information retrieval*. (Rapor No. ORNL/TM-13756) Computer Science and Mathematics Division, Oak Ridge National Laboratory. Oak Ridge, TN. Erişim adresi: <http://www.sandia.gov/~tgkolda/pubs/pubfiles/ornl-tm-13756.pdf>
- Clark, B. (2013). *Relevance theory*. Cambridge: Cambridge University Press.
- Cole, J. R. (2000). A short history of the use of citations as a measure of the impact of scientific and scholarly work. *The Web of Knowledge: A festschrift in honor of Eugene Garfield* içinde (s. 281-298). New Jersey: Information Today, Inc.
- Comins, J. A. ve Leydesdorff, L. (2016a). Identification of long-term concept-symbols among citations: Can documents be clustered in terms of common



intellectual histories? (*arXiv:1601.00288.*) Erişim adresi:  
<http://arxiv.org/abs/1601.00288>

Comins, J. A. ve Leydesdorff, L. (2016b). RPYS i/o: A web-based tool for the historiography and visualization of citation classics, sleeping beauties, and research fronts. (*arXiv:1602.01950.*) Erişim adresi:  
<http://arxiv.org/abs/1602.01950>

Cooper, W. S. (1978). Indexing documents by gedanken experimentation. *Journal of the American Society for Information Science*, 29(3), 107-119.

Cooper, W. S. ve Maron, M. E. (1978). Foundations of probabilistic and utility-theoretic indexing. *Journal of the ACM*, 25(1), 67–80.

Cortes, C. ve Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.

Croft, W. B. (2000). *Combining approaches to information retrieval*. W. B. Croft (Yay. haz.), *Advances in information retrieval: Recent research from the center for intelligent information retrieval* içinde (s.1-36). Boston: Kluwer.

Croft, W. B. ve Harper, D. J. (1979). Using probabilistic models of document retrieval without relevance information. *Journal of Documentation*, 35(4), 285-295.

Croft, W. B. ve Lafferty, J. (Yay. haz.). (2003). *Language modeling for information retrieval*. Boston: Kluwer Academic.

Çapkın, Ç. (2011) *Türkçe metin tabanlı açık arşivlerde kullanılan dizinleme yönteminin değerlendirilmesi*. (Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara). Erişim adresi:  
[http://eprints.rclis.org/28804/1/Cagdas\\_CAPKIN\\_Yuksele\\_Lisans\\_tezi.pdf](http://eprints.rclis.org/28804/1/Cagdas_CAPKIN_Yuksele_Lisans_tezi.pdf)

De Solla Price, D. (1965). Networks of scientific papers. *Science*, 149(3683), 510-515.

- Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. ve Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X. ve Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, 65(9), 1820-1833.
- Düzyol, G. (2011). *Türkiye kütüphanecilik ve bilgilim literatürünün entellektüel haritasının çıkarılması: Bir yazar ortak atıf analizi çalışması*. (Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara). Erişim adresi: [http://www.bby.hacettepe.edu.tr/akademik/guledaduzyoldogan/file/Guleda-Duzyol\\_Yukseklisans-Tezi.pdf](http://www.bby.hacettepe.edu.tr/akademik/guledaduzyoldogan/file/Guleda-Duzyol_Yukseklisans-Tezi.pdf)
- Emirzade, E. ve Bitirim, Y. (2008). Resim arama motorlarının sorgu sözcük sayısına göre performans değerlendirmeleri. *Bilgi Dünyası*, 9(1), 126-139. Erişim adresi: <http://bd.org.tr/index.php/bd/article/view/174/157>
- Fairthorne, R. (1961). *Towards information retrieval*. London: Butterworths.
- Fenton, J. E., Roy, D., Hughes, J. P. ve Jones, A. S. (2002). A century of citation classics in otolaryngology-head and neck surgery journals. *The Journal of Laryngology and Otology*, 116, 494-498.
- Fuhr, N. (1986). Two models of retrieval with probabilistic indexing. F. Rabitti (Yay. haz.), *ACM Conference on research and development in information retrieval (s. 249–257), 8–10 September 1986, Pisa, Italy*.
- Fuhr, N. (1989). Models for retrieval with probabilistic indexing. *Information Processing and Management*, 22(1), 55–72.
- Funkhouser, E. T. (1996). The evaluative use of citation analysis for communication journals. *Human Communication Research*, 22(4), 563-574.

- Galvez, C. ve Moya-Anegón, F. (2006). The unification of institutional addresses applying parametrized finite-state graphs (P-FSG). *Scientometrics*, 69(2), 323-345.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation - Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060), 471-479. doi: 10.1126/science.178.4060.471
- Glushko, R. J. (Yay. haz.). (2015). *The discipline of organizing: Professional edition* (3.bs) Sebastopol, CA.: O'Reilly Media.
- Good, I. J. (1958). *Speculations concerning information retrieval*. IBM Research Report RC-78, 10 Aralık 1958.
- Göker, A. ve Davies, J. (Yay. haz.). (2009). *Information retrieval: searching in the 21st century*. Chichester, UK: John Wiley & Sons.
- Harter, S. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*. 43(9), 602-615.
- Hersh, W. ve Voorhees, E. (2009). TREC genomics special issue overview. *Information Retrieval*, 12(1), 1-15.
- Hiemstra, D. (2000). A probabilistic justification for using tf×idf term weighting in information retrieval. *International Journal on Digital Libraries*, 3(2), 131-139.
- Jeong, Y. K., Song, M. ve Ding, Y. (2014). Content-based author co-citation analysis. *Journal of Informetrics*, 8(1), 197-211.
- Kahneman, D. (2011). *Hızlı ve Yavaş Düşünme*. İstanbul: Varlık Yayınları.
- Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. *Proceedings of the ACM-SIAM Symposium on discrete algorithms* içinde (s.1-33). Erişim adresi: <https://www.cs.cornell.edu/home/kleinber/auth.pdf>

- Larsen, B. (2004). *References and citations in automatic indexing and retrieval systems-experiments with the boomerang effect* (Doktora tezi, Kopenhag Üniversitesi). Erişim adresi: [http://pure.iva.dk/files/31034810/birger\\_larsen\\_phd.pdf](http://pure.iva.dk/files/31034810/birger_larsen_phd.pdf)
- Lin, X., White, H. D. ve Buzydlowski, J. (2003). Real-time author co-citation mapping for online searching. *Information Processing and Management*, 39, 689–706. Erişim adresi: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.4125&rep=rep1&type=pdf>
- Luhn, H. P. (1961). The automatic derivation of information retrieval encodements from machine-readable texts. A. Kent (Yay. haz). *Information retrieval and machine translation* içinde (Cilt 3, Bölüm 2, ss. 1021–1028). New York: Interscience Publication.
- Manning, C. ve Nayak, P. (2015). *Scoring, term weighting and the vector space model* [Ders slaytları]. Erişim adresi: <http://goo.gl/41iVs9>
- Manning, C., Raghavan, P. ve Schütze, H. (2008). *An introduction to information retrieval*. Cambridge: Cambridge University Press. Erişim adresi: <http://nlp.stanford.edu/IR-book/pdf/irbookprint.pdf>
- Manning, C. ve Schütze, H. (2000). *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press. Erişim adresi: [http://ics.upjs.sk/~pero/web/documents/pillar/Manning\\_Schuetze\\_Statistical\\_NLP.pdf](http://ics.upjs.sk/~pero/web/documents/pillar/Manning_Schuetze_Statistical_NLP.pdf)
- Maron, M. E. (1977). On indexing, retrieval and the meaning of about. *Journal of the American Society for Information Science*, 28(1), 38–43.
- Maron, M. E. (1984). Probabilistic retrieval models. B. Dervin ve M. J. Voigt (Yay. haz), *Progress in communication sciences* içinde (Vol. 5, pp. 145–176).

- Maron, M. E. ve Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. *Journal of the ACM*, 7, 216-244.
- Maron, M. E., Curry, S. ve Thompson, P. (1986). An inductive search system: Theory, design and implementation. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16(1), 21–28.
- Maron, M. E., Kuhns, J. L ve Ray, L. (1958). *Some experiments with probabilistic indexing*. Los Angeles, CA: Thomson Ramo-Wooldridge Inc.
- Mayr, P. ve Scharnhorst, A. (2015). Scientometrics and information retrieval: weak-links revitalized. *Scientometrics*, 102(3), 2193-2199.
- McHale, M. (1998). A comparison of WordNet and Roget's taxonomy for measuring semantic similarity. *Proceedings of the workshop on usage of WordNet in natural language processing systems* içinde (s. 115–120). Erişim adresi: <http://www.aclweb.org/anthology/W98-0716>
- McNee, S. M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S. K., Rashid, A. M. ve Riedl, J. (2002). On the recommending of citations for research papers. *Proceedings of the 2002 ACM Conference on computer supported cooperative work* içinde (s. 116-125). Erişim adresi: <http://www-users.cs.umn.edu/~mcnee/mcnee-cscw2002.pdf>
- Merton, R. K. (1968). The Matthew effect in science. *Science*. 159(3810), 56–63.
- Merton, R. K. (1988). The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property. *ISIS*, 79, 606-623.
- Miller (1971). A probabilistic search strategy for Medlars. *Journal of Documentation*, 27, 254–266. Erişim adresi: <http://www.ictir2013.org/miller.pdf>
- Papp, I., Ercsey-Ravasz, M., Deritei, D., Sumi, R., Járai-Szabó, F., Florian, R.V. ve diğerleri (2013). The P-Index: Hirsch Index of Individual Publications.

*Proceedings of ISSI 2013 Vienna* içinde (s. 2086-2088). Vienna, Austria, July 15-19, 2013.

Poetz, M. K. ve Prügl, R. (2010). Crossing domain-specific boundaries in search of innovation: Exploring the potential of pyramiding. *Journal of Product Innovation Management*, 27(6), 897-914.

Prevedelli, D., Simonini, R. ve Ansaloni, I. (2001). Relationship of non-specific commensalism in the colonization of the deep layers of sediment. *Journal of the Marine Biological Association of the UK*, 81, 897-901. doi:10.1017/S0025315401004817.

Robertson, S. E. (2003). The unified model revisited. *Workshop on the mathematical foundations of information retrieval, SIGIR 2003, Toronto, Canada*. Erişim adresi: [http://www.staff.city.ac.uk/~sb317/papers/UM\\_revisited.pdf](http://www.staff.city.ac.uk/~sb317/papers/UM_revisited.pdf)

Robertson, S. E. (1977). The probability ranking principle in IR. *Journal of Documentation*, 33(4), 294–304. Erişim adresi: <http://parnec.nuaa.edu.cn/xtan/IIR/readings/jdRobertson1977.pdf>

Robertson, S. E., Maron, M. E. ve Cooper, W. S. (1982). Probability of relevance: A unification of two competing models for document retrieval. *Information Technology: Research and Development*, 1, 1–21.

Robertson, S. E., Maron, M. E., ve Cooper, W. S. (1983). The unified probabilistic model for IR. *Research and development in information retrieval* içinde (s. 108-117). Berlin, Heidelberg: Springer.

Robertson, S. E. ve Sparck Jones, K. (1976). Relevance weighting of search terms. *Journal of the American Society for Information Science*, 27(3), 129-146.

Rocchio, J. J. (1966). *Document retrieval systems: Optimization and evaluation*. (Doktora Tezi, Harvard Üniversitesi), Rapor No. ISR-10 to the National Science Foundation, Harvard Computation Laboratory.

- Salton, G. ve Lesk, M. E. (1968). Computer evaluation of indexing and text processing. *Journal of the ACM*, 15, 8–36.
- Salton, G. ve McGill, M. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Salton, G., Wong, A. ve Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18 (11), 613-620.
- Salton, G. ve Yang, C. S. (1973). On the specification of term values in automatic indexing. *Journal of Documentation*, 29, 351–372.
- Saracevic, T. (1975). Relevance: A review of and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science*, 26, 321–343.
- Saracevic, T. (1996). Relevance reconsidered. *Proceedings of the second conference on conceptions of library and information science* içinde (s. 201-218). Lokve, Croatia: Benja Publishing.
- Saracevic, T. ve Kantor, P. (1988a). A study of information seeking and retrieving. II. Users, questions and effectiveness. III. Searchers, searches and overlap. *Journal of the American Society for Information Science*, 39(3), 177–196.
- Saracevic, T. ve Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches and overlap. *Journal of the American Society for Information Science*, 39(3), 197–216.
- Saracevic, T., Kantor, P., Chamis, A. Y. ve Trivison, D. (1988). A study of information seeking and retrieving. I. Background and methodology. *Journal of the American Society for Information Science*, 39(3), 161–176.
- Shah, C. (Yapımcı). (2009, 6 Ekim). *INLS490-154W: Information retrieval systems design and implementation* [Podcast]. Erişim adresi:

<https://itunes.apple.com/us/podcast/inls490-154w-information-retrieval/id327757721?mt=2>

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379-423.

Shannon, C. E. ve Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Ill: University of Illinois Press.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, 24(4), 265-269.

Smucker, M. D. (2008) *Evaluation of find-similar with simulation and network analysis*. (Doktora Tezi, University of Massachusetts, Amherst). Erişim adresi: <http://www.mansci.uwaterloo.ca/~msmucker/publications/smucker-dissertation.pdf>

Soll, J. B., Milkman, K. L., ve Payne, J. W. (2015). Outsmart your own biases. *Harvard Business Review*, 93(5), 64-71.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28(1): 11–21. doi: 10.1108/eb026526.

Sparck Jones, K. (1973). Index term weighting. *Information Storage and Retrieval*, 9(11), 619-633.

Sperber, D. ve Wilson, D. (1995). *Relevance: Communication and cognition*. Oxford: Blackwell. Erişim adresi: [http://monoskop.org/images/e/e6/Sperber\\_Dan\\_Wilson\\_Deirdre\\_Relevance\\_Communications\\_and\\_Cognition\\_2nd\\_edition\\_1996.pdf](http://monoskop.org/images/e/e6/Sperber_Dan_Wilson_Deirdre_Relevance_Communications_and_Cognition_2nd_edition_1996.pdf)

Stanford. (2012, Ocak). Jurafsky, D. ve Manning, C.: *Ranked information retrieval* [Çevrimiçi ders videosu]. Erişim adresi: <https://class.coursera.org/nlp/lecture/179>



- Strohman, T., Croft, W. B. ve Jensen, D. (2007). Recommending citations for academic papers. *Proceedings of the 30th Annual International ACM SIGIR Conference on research and development in information retrieval* içinde (s. 705-706). Amsterdam, The Netherlands: ACM.
- Sylvia, M. J. (1998). Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies. *Collection Building*, 17(1), 20-28.
- Taşkın, Z. (2012). *Atıf dizinlerinde üniversite adreslerinin standardizasyon sorunu*. (Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara). Erişim adresi: [http://www.bby.hacettepe.edu.tr/yayinlar/dosyalar/thesis\\_zt.pdf](http://www.bby.hacettepe.edu.tr/yayinlar/dosyalar/thesis_zt.pdf)
- Thompson, P. (1988). Subjective probability and information retrieval: A review of the psychological literature. *Journal of Documentation*, 44(2), 119–143.
- Thompson, P. (1990a). A combination of expert opinion approach to probabilistic information retrieval, Part 1: The conceptual model. *Information Processing and Management*, 26(3), 371–382.
- Thompson, P. (1990b). A combination of expert opinion approach to probabilistic information retrieval, Part 2: Mathematical treatment of CEO model 3. *Information Processing and Management*, 26(3), 383–394.
- Thompson, P. (2007). Looking back: On relevance, probabilistic indexing and information retrieval. *Information Processing and Management*, 44(2), 963-970.
- Thomson Reuters (Yapımcı). (2015, 26 Mayıs). *The Importance of bibliometrics in research evaluation* [Podcast]. Erişim adresi: <https://itunes.apple.com/us/podcast/mena-talks-ip-science/id992815770?mt=2>
- Tonta, Y. (1995). Bilgi erişim sistemleri. *Türk Kütüphaneciliği*, 9(3): 302-314.

- Tonta, Y. (2001). Bilgi erişim sorunu. T. Fenerci ve O. Gürdal (Yay. haz), 21. *Yüzyıla Girerken Enformasyon Olgusu: Ulusal Sempozyum Bildirileri, 19-20 Nisan 2001, Hatay* içinde (s. 198-206). Ankara: Türk Kütüphaneciler Derneği, 2001.
- Tonta, Y. (2002). Bilgi erişim sorunları ve Internet. Ali Can, M. T. Gülle, O. Gürdal ve E. Yılmaz (Yay. haz.). *Kütüphanecilikte Yeni Gelişmeler, Kavramlar, Olgular... 37. Kütüphane Haftası bildirileri. 26 Mart - 01 Nisan 2001* içinde (s. 52-62). Ankara: TKD, 2002.
- Tonta, Y., Bitirim, Y. ve Sever, H. (2002). *Türkçe arama motorlarında performans değerlendirme*. Ankara: Total Bilişim. Erişim adresi: <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-bitirim-sever-arama-motorlari.pdf>
- Tonta, Y. (2012). Bilgi sınıflama, bilgi düzenleme ve bilgi erişim. Ö. Külcü, T. Çakmak ve N. Özel (Yay. Haz.). *Prof. Dr. K. Gülbün Baydur'a armağan* içinde (s. 155-172). Ankara: Hacettepe Üniversitesi Bilgi ve Belge Yönetimi Bölümü, 2012. Erişim adresi: <http://yunus.hacettepe.edu.tr/~tonta/yayinlar/tonta-bilgi-siniflama-duzenleme-erisim.pdf>
- Tonta, Y. ve Özkan Çelik, A. E. (2013). Cahit Arf: Exploring his scientific influence using social network analysis, author co-citation maps and single publication h index. *Journal of Scientometric Research*, 2:37-51.
- Turtle, H. R ve Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. *ACM Transactions on Information Systems*, 9(3), 187–222.
- Turtle, H. R. ve Croft, W. B. (1997). Uncertainty in information retrieval systems. A. Motro, P. Smets (Yay. Haz.). *Uncertainty management in information systems: From needs to solutions* içinde (s. 189-224). Boston: Kluwer Academic.

- Tversky, A. ve Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), s. 1124-1131.
- Uçak, N. Ö. ve Al, U. (2009). Bilimsel iletişimin zamana göre değişimi: Bir atıf analizi çalışması. *Bilgi Dünyası*, 10(1), 1-22.
- Van Rijsbergen, C. J. (2005). The emergence of probabilistic accounts of information retrieval. J. Tait (Yay. haz.), *Charting a new course: Natural language processing and information retrieval essays in honour of Karen Sparck Jones* içinde (s. 23–38). Dordrecht, The Netherlands: Springer.
- Van Rijsbergen, C. J., Harper, D. J. ve Porter, M. F. (1981). The selection of good search terms. *Information Processing and Management*, 17(2), 77-91.
- Wang, J. (2014). Unpacking the Matthew effect in citations. *Journal of Informetrics*, 8(2), 329-339.
- Web of Science (2010). Related records. Erişim adresi: [https://images.webofknowledge.com/WOK48B3/help/WOS/hp\\_related\\_records.html](https://images.webofknowledge.com/WOK48B3/help/WOS/hp_related_records.html)
- White, H. D. (2003). Pathfinder networks and author co-citation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology*, 54, 423-434.
- White, H. D. (2007a). Combining bibliometrics, information retrieval, and relevance theory. Part 1: First examples of a synthesis. *Journal of the American Society for Information Science and Technology*, 58, 536-559.
- White, H. D. (2007b). Combining bibliometrics, information retrieval, and relevance theory. Part 2: Some implications for information science. *Journal of the American Society for Information Science and Technology*, 58, 583-605.
- White, H. D. (2009). Pennants for Strindberg and Persson. *Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th*

*birthday*. Special volume of the *E-newsletter of the International Society for Scientometrics and Informetrics*, 5, 71-83.

White, H. D. (2010). Some new tests of relevance theory in information science. *Scientometrics*, 83, 653–667.

White, H. D. (2011). Relevance theory and citations. *Journal of Pragmatics*, 43, 3345–3361. Erişim adresi: [http://www.researchgate.net/publication/229861357\\_Relevance\\_theory\\_and\\_citations](http://www.researchgate.net/publication/229861357_Relevance_theory_and_citations)

White, H. D. (2015). Co-cited author retrieval and relevance theory: examples from the humanities. *Scientometrics*, 102(3), 2275-2299.

White, H. D. (2016). Bag of works retrieval: TF\*IDF weighting of co-cited works. *Proceedings of the 3rd workshop on bibliometric-enhanced information retrieval (BIR2016)* içinde (s. 63-72). Erişim adresi: <http://ceur-ws.org/Vol-1567/paper7.pdf>

White, H. D. ve Griffith, B. C. (1981). Author co-citation: a literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163-171.

White, H. D. ve McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-355.

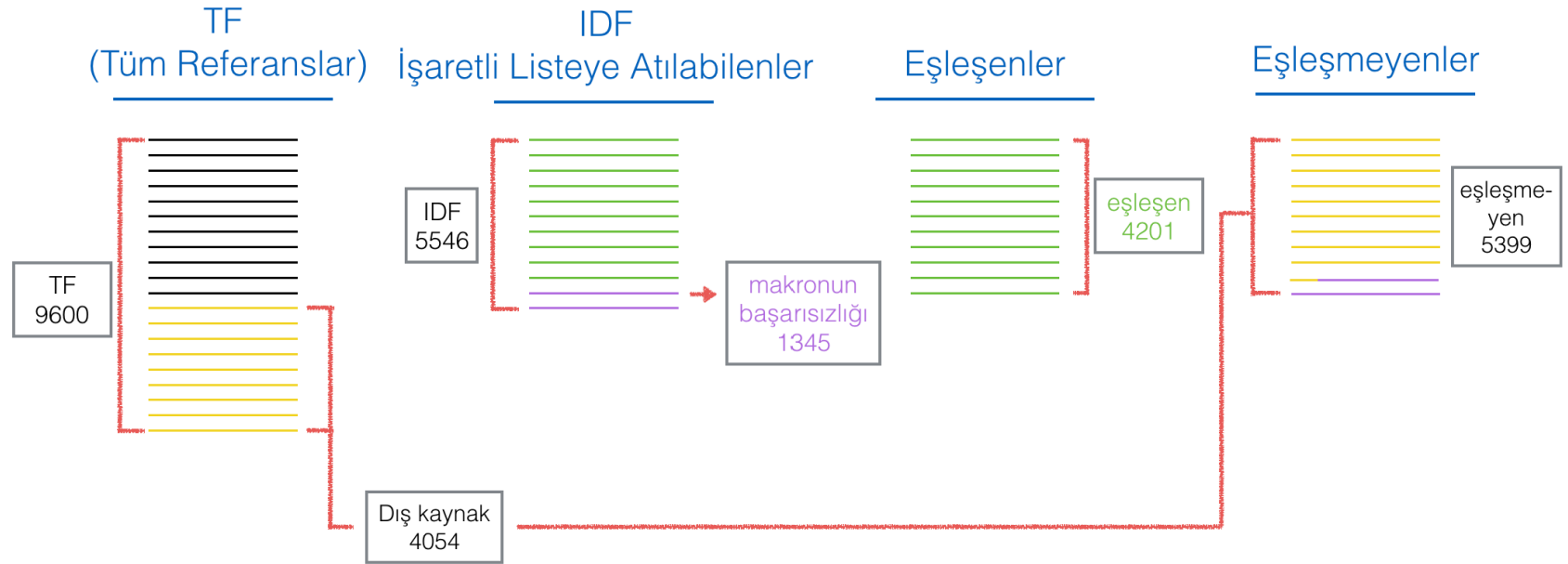
White, H. D. ve Mayr, P. (2013). Pennants for descriptors. *NKOS Workshop 2013. Valletta, Malta*. Erişim adresi: <http://arxiv.org/abs/1310.3808>

White, H. D., Buzydlowski, J. ve Lin, X. (2000) Co-cited author maps as interfaces to digital libraries: Designing pathfinder networks in the humanities. *Proceedings of the International Conference on Information Visualization* (s. 25–30). London: Computer Society Press.

- White, H. D., Lin, X. ve Buzydlowski, J. W. (2001). The endless gallery: Visualizing authors' citation images in the humanities. *Proceedings of the American Society of Information Science and Technology Annual Meeting* içinde (s. 182–189). Washington, DC: ASIS&T.
- Wilson, P. (1973). Situational relevance. *Information Storage and Retrieval*, 9, 457-471.
- Wilson, D. ve Sperber, D. (2002). Relevance theory. G. Ward ve L. Horn (Yay. haz.) *Handbook of pragmatics* içinde. Oxford: Blackwell. Erişim adresi: [http://cogprints.org/2317/1/relevance\\_theory.htm#\\_edn1](http://cogprints.org/2317/1/relevance_theory.htm#_edn1)
- Wolfram, D. (2000). Applications of informetrics to information retrieval research. *Informing Science*, 3(2): 77-82.
- Wolfram, D. (2015). The symbiotic relationship between information retrieval and informetrics. *Scientometrics*, 102(3), 2201-2214.
- Wu, H. C., Luk, R. W. P., Wong, K. F. ve Kwok, K. L. (2008). Interpreting tf\*idf term weights as making relevance decisions. *ACM Transactions on Information Systems (TOIS)*, 26(3), 1-37.
- Yıldırım, B. F. ve Yeşilyurt, C. (2014). Bulanık analitik hiyerarşi prosesi yaklaşımı ile proje değerlendirme kriterlerinin önceliklendirilmesi: Kalkınma ajansı örneği. *Atatürk İletişim Dergisi*, 6(1), 25-53. ISSN: 2146-1538.
- Yule, G. U. (1912). On the methods of measuring association between two attributes. *Journal of the Royal Statistical Society*, 75(6), 579-652.
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.
- Zhang, G., Ding, Y. ve Milojević, S. (2013). Citation content analysis (cca): A framework for syntactic and semantic analysis of citation content. *Journal of the American Society for Information Science and Technology*, 64(7), 1490-1503.

## EKLER

### Ek 1: Pennant diyagramının hazırlanması için uygulanan işlemler



Ortak atıf saydırma makrosu,  
birikimli hale getirme makrosu

İşaretili listeden oluşturulan atıf raporu  
sonuçlarını çekme,  
formata uydurma ve birikimli hale getirme

Birleştirme (yazar, yıl, tam dergi adı ve kısaltması eşleştirmesi),  
atıfları birikimli hale getirme ve  
Pennant formatı haline getirme makroları

## Ek 2: Örnekleme dayanan pennant diyagramlarında (Şekil 7 ve 12) temsil edilen çalışmalar

Sıra No	Künye	Konu	Sektör/Alan
(37)	Fuhr, N. (1986). Two models of retrieval with probabilistic indexing. F. Rabitti (Yay. haz.), <i>ACM Conference on research and development in information retrieval</i> (s. 249–257), 8–10 September 1986, Pisa, Italy.	olasılıksal bilgiye erişim	A/6-1
(39)	Bodoff, D. (1999). A re-unification of two competing models for document retrieval. <i>Journal of the American Society for Information Science</i> , 50(1), 49–64.	model 1 ve model 3	A/6-1
(44)	Luk, R. W. (2008). On event space and rank equivalence between probabilistic retrieval models. <i>Information Retrieval</i> , 11(6), 539-561.	indekslemede bilgisayar desteği	A/6-1
(46)	Maron, M. E. (1984). Probabilistic retrieval models. B. Dervin ve M. J. Voigt (Yay. haz.), <i>Progress in communication sciences</i> (s. 145–176) içinde.	arama stratejisi	A/6-1
(54)	Good, I. J. (1958). <i>Speculations concerning information retrieval</i> . IBM Research Report RC-78, 10 Aralık 1958.	model 3	A/6-1
(57)	Robertson, S. E. (2003). The unified model revisited. <i>Workshop on the mathematical foundations of information retrieval, SIGIR 2003, Toronto, Canada</i> .	model 3	A/6-1
(58)	Robertson, S. E., Maron, M. E., ve Cooper, W. S. (1983). The unified probabilistic model for IR. <i>Research and development in information retrieval</i> (s. 108-117) içinde. Berlin, Heidelberg: Springer.	olasılıksal indeksleme	C/6-1
(59)	Hersh, W. ve Voorhees, E. (2009). TREC genomics special issue overview. <i>Information Retrieval</i> , 12(1), 1-15.	makine öğrenmesi, deneysel çalışmalar	A/6-1
(61)	Hodge, G. M. ve Milstead, J. L. (1998). <i>Computer support to indexing</i> . Philadelphia, PA: National Federation of Abstracting and Information Services.	model 1, model 3	A/6-1
(63)	Cleverdon, C. W. (1959). The evaluation of systems used in information retrieval. <i>Proceedings of the international conference on scientific information</i> (s. 687-698) içinde. Washington, DC: National Academy of Sciences	olasılıksal model, sıralama	A/6-1
(64)	Fairthorne, R. (1961). <i>Towards information retrieval</i> . London: Butterworth.	olasılıksal erişim modelleri	A/6-1
(65)	Radecki, T. (1976). Mathematical model of information retrieval system based on the concept of fuzzy thesaurus. <i>Information Processing and Management</i> , 12(5), 313-318.	olasılıksal konseptler, istatistiksel yöntemler, bulanık mantık	A/6-1
Sıra No	Künye	Konu	Sektör/Alan
(9)	Winkler, R. L. (1968). The consensus of subjective probability distributions. <i>Management Science</i> , 15(2), B-61- B-75.	olasılık dağılımı, bayes	A/5-1
(21)	Miller (1971). A probabilistic search strategy for Medlars. <i>Journal of Documentation</i> , 27, 254–266.	model 2	A/5-1
(22)	Thompson, P. (1988). Subjective probability and information retrieval: A review of the psychological literature. <i>Journal of Documentation</i> , 44(2), 119–143.	subjektif olasılık	A/5-1
(23)	Bodoff, D. ve Robertson, S. E. (2004). New unified probabilistic model. <i>Journal of the American Society for Information Science and Technology</i> , 55(6), 471–487.	model 1 ve model 3	A/5-1
(28)	Maron, M. E., Curry, S. ve Thompson, P. (1986). An inductive search system: Theory, design and implementation. <i>IEEE Transactions on Systems, Man, and Cybernetics, SMC-16(1)</i> , 21–28	model 2	A/5-1
(32)	Dos Santos, E. B., Hruschka E. R., Hruschka, E. R. ve Ebecken, N. F. (2011). Bayesian network classifiers: Beyond classification accuracy. <i>Intelligent Data Analysis</i> , 15(3), 279-298.	bayes, sınıflama	A/5-1
(36)	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, Part 1: The conceptual model. <i>Information Processing and Management</i> , 26(3), 371–382.	model 3	A/5-1
(40)	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, Part 2: Mathematical treatment of CEO model 3. <i>Information Processing and Management</i> , 26(3), 383–394.	model 3	A/5-1
(50)	Croft, W. B. (2000). Combining approaches to information retrieval. W. B. Croft (Yay. haz.), <i>Advances in information retrieval: Recent research from the center for intelligent information retrieval</i> (s.1-36) içinde. Boston: Kluwer.	olasılıksal sıralama	A/5-1
(53)	Rocchio, J. J. (1966). <i>Document retrieval systems: Optimization and evaluation</i> . (Doktora Tezi, Harvard Üniversitesi), Rapor No. ISR-10 to the National Science Foundation, Harvard Computation Laboratory.	ilgiliilik geribildirimi	A/5-1
(67)	Bagley, P. (1951). <i>Electronic digital machines for high-speed information searching</i> . (Yüksek Lisans Tezi, Massachusetts Teknoloji Enstitüsü, Cambridge).	belge erişim sistemi	A/5-1
(68)	Van Rijsbergen, C. J. (2005). The emergence of probabilistic accounts of information retrieval. J. Tait (Yay. haz.), <i>Charting a new course: Natural language processing and information retrieval essays in honour of Karen Sparck Jones</i> (s. 23–38) içinde. Dordrecht, The Netherlands: Springer.	model 1 ve model 2	A/5-1

Sıra No	Künye	Konu	Sektör/Alan
(25)	Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. <i>Annual Review of Information Science and Technology</i> , 20, 117-151.	olasılık ve bulanık küme	A/4-1
(27)	Wong, S. K. M. ve Yao, Y. Y. (1995). On modeling information retrieval with probabilistic inference. <i>ACM Transactions on Information Systems (TOIS)</i> , 13(1), 38-68.	olasılıksal çıkarım	A/4-1
(33)	Gordon, M. (1988). Probabilistic and genetic algorithms in document retrieval. <i>Communications of the ACM</i> , 31(10), 1208-1218.	olasılık, genetik algoritmalar	B/4-1
(34)	Swanson, D. R. (1977). Information retrieval as a trial-and-error process. <i>The Library Quarterly</i> , 128-148.	bilgi erişim	B/4-1
(38)	Van Rijsbergen, C. J., Harper, D. J. ve Porter, M. F. (1981). The selection of good search terms. <i>Information Processing and Management</i> , 17(2), 77-91.	kullanıcı sorguları	B/4-1
(41)	Cleverdon, C.W. (1962). <i>Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems</i> . Cranfield College of Aeronautics,.	manuel indeksleme şeması	B/4-1
(43)	Salton, G. ve Lesk, M. E. (1968). Computer evaluation of indexing and text processing. <i>Journal of the ACM</i> , 15, 8-36.	indeksleme, deneysel çalışma	B/4-1
(45)	Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. <i>Proceedings of the ACM-SIAM Symposium on discrete algorithms</i> (s. 1-33) içinde.	hiperbağlantı, HITS grafik yapılı algoritma	B/4-1
(51)	Saracevic, T. ve Kantor, P. (1988b). A study of information seeking and retrieving. III. Searchers, searches and overlap. <i>Journal of the American Society for Information Science</i> , 39(3), 197-216.	olasılıksal sıralama	B/4-1
(55)	Ward J. H. ve Hook, M. E. (1963). Application of an Hierarchical Grouping Procedure to a Problem of Grouping Profiles. <i>Educational and Psychological Measurement</i> , 23, 69-81.	hiyerarşik gruplama	B/4-1
(60)	Brin, S. ve Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. <i>Computer Networks and ISDN Systems</i> , 30(1-7), 107-117. doi: 10.1016/s0169-7552(98)00110-x	page rank	B/4-1
(74)	Kleinberg, J. M. (1998). Decision algorithms for unsplitable flow and the half-disjoint paths problem. <i>Proceedings of the thirtieth annual ACM symposium on Theory of computing</i> içinde (s. 530-539). Dallas, Texas, USA: ACM.	karar algoritmaları	B/4-1
Sıra No	Künye	Konu	Sektör/Alan
(49)	Chow, C. ve Liu, C. (1968). Approximating discrete probability distributions with dependence trees. <i>IEEE Transactions on Information Theory</i> , 14(3), 462-467.	olasılık dağılımları	C/3-1
(52)	Tague, J. M. (1981). User-responsive subject control in bibliographic retrieval systems. <i>Information Processing and Management</i> , 17(3), 149-156.	bilbiyografik erişim sistemleri	C/3-1
(62)	Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. <i>Journal of the American Society for information Science</i> , 24(4), 265-269.	bibliyometri	C/3-1
(66)	De Solla Price, D. (1965). Networks of scientific papers. <i>Science</i> , 149(3683), 510-515.	bibliyometri	C/3-1
(67)	Chen, M. S., Han, J. ve Yu, P. S. (1996). Data mining: An overview from a database perspective. <i>IEEE Transactions on Knowledge and Data Engineering</i> , 8(6), 866-883.	veri madenciliği	C/3-1
(70)	Goodman, L. A. ve Kruskal, W. H. (1979). Measures of association for cross classifications. <i>Measures of association for cross classifications</i> (s. 2-34) içinde. New York: Springer Verlag New York Inc.	sınıflama	C/3-1
(71)	Frei, H. P. ve Stieger, D. (1995). The use of semantic links in hypertext information retrieval. <i>Information Processing and Management</i> , 31(1), 1-13.	bilgi erişim, olasılık	C/3-1
(72)	Garfield, E. (1972). Citation analysis as a tool in journal evaluation - Journals can be ranked by frequency and impact of citations for science policy studies. <i>Science</i> , 178(4060), 471-479. doi: 10.1126/science.178.4060.471	atf analizi, sıralama	C/3-1
(76)	Markey, K. (1981). Levels of question formulation in negotiation of information need during the online pre-search interview: A proposed model. <i>Information Processing and Management</i> , 17(5), 215-225.	bilgi gereksinimi	C/3-1
(81)	Salton, G. ve McGill, M. (1983). <i>Introduction to modern information retrieval</i> . New York: McGraw-Hill.	bilgi erişim	C/3-1
(82)	Wooldridge, M. ve Jennings, N. R. (1995). Intelligent agents: Theory and practice. <i>The Knowledge Engineering Review</i> , 10(02), 115-152.	akıllı ajanlar	C/3-1
(83)	Losee, R. M. (1988). Parameter estimation for probabilistic document-retrieval models. <i>Journal of the American Society for Information Science</i> , 39(1), 8-16.	olasılıksal modeller	C/3-1



Sıra No	Künye	Konu	Sektör/Alan
(75)	Dempster, A. P., Laird, N. M. ve Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. <i>Journal of the Royal Statistical Society. Series B (methodological)</i> , 1-38.	maksimum olabilirlik	C/2-1
(77)	Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. <i>Proceedings of the IEEE</i> , 77(2), 257-286.	bilimsel çalışmaların ağları	C/2-1
(78)	Shannon, C. E. ve Weaver, W. (1949). <i>The mathematical theory of communication</i> . Urbana, Ill: University of Illinois Press.	matematiksel iletişim kuramı	C/2-1
(79)	Freund, Y. ve Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. In <i>European Conference on computational learning theory</i> (s. 23-37) içinde. Heidelberg, Berlin: Springer.	karar teorik, boosting algoritması	C/2-1
(80)	Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. <i>Journal of the American Statistical Association</i> , 58(301), 236-244.	hiyerarşik gruplama	C/2-1
(84)	Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. <i>Proceedings of the national academy of sciences</i> , 79(8), 2554-2558.	sinir ağları	C/2-1
(85)	Cortes, C. ve Vapnik, V. (1995). Support-vector networks. <i>Machine Learning</i> , 20(3), 273-297.	örüntü tanıma, sinir ağları, OCR	C/2-1
(86)	Breiman, L. (2001). Random forests. <i>Machine Learning</i> , 45(1), 5-32.	Karar ağacı, rastgele orman	C/2-1
(87)	Tversky, A. ve Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. <i>Science</i> , 185(4157), s. 1124-1131.	karar verme, belirsizlik, önyargı	C/2-1
(88)	Watts, D. J. ve Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. <i>Nature</i> , 393(6684), 440-442.	küçük dünya ağları	C/2-1
(89)	Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-III. <i>Information Sciences</i> , 9(1), 43-80.	kestirimsel akıl yürütme	C/2-1
(90)	Akaike, H. (1974). A new look at the statistical model identification. <i>IEEE Transactions on Automatic Control</i> , 19(6), 716-723.	dil modeli	C/2-1

Sıra No	Künye	Konu	Sektör/Alan
(1)	Cooper, W. S. ve Maron, M. E. (1978). Foundations of probabilistic and utility-theoretic indexing. <i>Journal of the ACM</i> , 25(1), 67-80.	subjektif yararlılık, bilgi erişim sistemleri, matematiksel model	A/5-2
(6)	Robertson, S. E. (1994). Documentation note query-document symmetry and dual models. <i>Journal of Documentation</i> , 50(3), 233-238.	doküman sorgu simetrisi	A/5-2
(7)	Salton, G., Fox, E. A. ve Wu, H. (1983). Extended Boolean information retrieval. <i>Communications of the ACM</i> , 26(11), 1022-1036.	genişletilmiş boole	A/5-2
(8)	Bookstein, A. ve Kraft, D. (1977). Operations research applied to document indexing and retrieval decisions. <i>Journal of the ACM</i> , 24(3), 418-427.	ilgiliilik	A/5-2
(10)	Cooper, W. S. (1978). Indexing documents by gedanken experimentation. <i>Journal of the American Society for Information Science</i> , 29(3), 107-119.	olasılıksal bilgi erişim	A/5-2
(11)	Bookstein, A. (1983). Outline of a general probabilistic retrieval model. <i>Journal of Documentation</i> , 39(2), 63-72.	indeksleme	A/5-2
(12)	Bookstein, A. ve Cooper, W. (1976). A general mathematical model for information retrieval systems. <i>The Library Quarterly</i> , 153-167.	makine indeksleme	A/5-2
(13)	Luhn, H. P. (1961). The automatic derivation of information retrieval encodements from machine-readable texts. A. Kent (Yay. haz). <i>Information retrieval and machine translation</i> (Cilt 3, Bölüm 2, ss. 1021-1028) içinde. New York: Interscience Publication.	indeksleme	A/5-2
(15)	Bookstein, A. (1978). On the perils of merging Boolean and weighted retrieval systems. <i>Journal of the American Society for Information Science</i> , 29(3), 156-158.	dil modeli	A/5-2
(16)	Croft, W. B. ve Lafferty, J. (Yay. haz.). (2003). <i>Language modeling for information retrieval</i> . Boston: Kluwer Academic.	sıralama, deneysel indeksleme, fayda kestirimi	A/5-2
(17)	Bookstein, A. (1979). Relevance. <i>Journal of the American Society for Information Science</i> , 30(5), 269-273.	boole, ağırlıklandırma	A/5-2
(18)	Swanson, D. R. (1960). Searching natural language text by computer. <i>Science</i> , 132(3434), 1099-1104.	ilgiliilik, fayda-teorik indeksleme, subjektif yararlılık	A/5-2

4-2: (42 kaynak)	Sıra No	Künye	Konu	Sektör/Alan
	(2)	Robertson, S. E., Maron, M. E. ve Cooper, W. S. (1982). Probability of relevance: A unification of two competing models for document retrieval. <i>Information Technology: Research and Development</i> , 1, 1–21.	model 1 ve model 2	A/4-2
	(3)	Robertson, S. E. (1977). The probability ranking principle in IR. <i>Journal of Documentation</i> , 33(4), 294–304.	olasılıksal sıralama	B/4-2
	(5)	Croft, W. B. ve Harper, D. J. (1979). Using probabilistic models of document retrieval without relevance information. <i>Journal of Documentation</i> , 35(4), 285-295.	olasılıksal model	B/4-2
	(14)	Fuhr, N. (1989). Models for retrieval with probabilistic indexing. <i>Information Processing and Management</i> , 22(1), 55–72.	olasılıksal indeksleme	A/4-2
	(19)	Maron, M. E. (1977). On indexing, retrieval and the meaning of about. <i>Journal of the American Society for Information Science</i> , 28(1), 38–43.	olasılıksal sıralama, hakkındalık	A/4-2
	(20)	Turtle, H. R ve Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. <i>ACM Transactions on Information Systems</i> , 9(3), 187–222.	ağ tabanlı erişim modelleri, vektör uzayı, sıralama, boole	B/4-2
	(24)	Salton, G., Yang, C. S. ve Yu, C. T. (1975). A theory of term importance in automatic text analysis. <i>Journal of the American Society for Information Science</i> , 26(1), 33-44.	olasılık, terim ağırlıklandırma, discrimination value analysis	B/4-2
	(26)	Saracevic, T. (1975). Relevance: A review of and a framework for thinking on the notion in information science. <i>Journal of the American Society for Information Science</i> , 26, 321–343.	ilgiliilik, sıralama	B/4-2
	(29)	Bush, V. (1945). As we may think. <i>Atlantic Monthly</i> , 176(1), 101-108.	memex, hypertext, ilişkili dizinleme	B/4-2
	(30)	Blair, D. C. ve Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. <i>Communications of the ACM</i> , 28(3), 289–299.	performans değerlendirme, tam metin bilgi erişim	B/4-2
	(31)	Salton, G. ve Yang, C. S. (1973). On the specification of term values in automatic indexing. <i>Journal of Documentation</i> , 29, 351–372.	otomatik indeksleme, ağırlıklandırma	B/4-2
	(35)	Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. <i>Journal of Documentation</i> 28(1): 11–21. doi: 10.1108/eb026526.	ters doküman sıklığı, tf*idf	B/4-2
3-2: (5 kaynak)	Sıra No	Künye	Konu	Sektör/Alan
	(4)	Robertson, S. E. ve Sparck Jones, K. (1976). Relevance weighting of search terms. <i>Journal of the American Society for Information Science</i> , 27(3), 129-146.	model 2	C/3-2
	(42)	Salton, G., Wong, A. ve Yang, C. S. (1975). A vector space model for automatic indexing. <i>Communications of the ACM</i> , 18 (11), 613-620.	vektör uzayı	C/3-2
	(47)	Porter, M. F. (1980). An algorithm for suffix stripping. <i>Program (Automated Library and Information Systems)</i> , 14(3), 130-137.	algoritmalar	C/3-2
	(56)	Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. ve Harshman, R. (1990). Indexing by latent semantic analysis. <i>Journal of the American Society for Information Science</i> , 41(6), 391.	gizli anlam indeksleme	C/3-2
(48)	Salton, G., Wong, A. ve Yu, C. T. (1976). Automatic indexing using term discrimination and term precision measurements. <i>Information Processing and Management</i> , 12(1), 43-51.	otomatik indeksleme	C/3-2	
2-2: (1 kaynak)	Sıra No	Künye	Konu	Sektör/Alan
(73)	Zadeh, L. A. (1965). Fuzzy sets. <i>Information and Control</i> , 8(3), 338-353.	bulanık küme	C/2-2	
4-3: (1 kaynak)	Sıra No	Künye	Konu	Sektör/Alan
(-)	Maron, M. E. ve Kuhns, J. L. (1960). On relevance, probabilistic indexing and information retrieval. <i>Journal of the ACM</i> , 7, 216-244.	olasılıksal dizinleme, ilgiliilik	C/4-3	


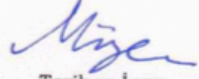
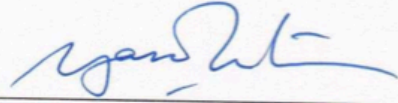
### EK 3 : WoS Kaynakça benzerliğine göre oluşturulan ilgililik sıralaması ile Pennant Diyagramı yöntemi ile oluşturulan ilgililik sıralamalarının karşılaştırılması

Sıra No	WoS Kaynakça Benzerliğine Göre Oluşturulan İlgililik Sıralaması	Pennant Diyagramları Yöntemine Göre Oluşturulan İlgililik Sıralaması
1	Prevedelli, D., Simonini, R. ve Ansaloni, I. (2001). Relationship of non-specific commensalism in the colonization of the deep layers of sediment. <i>Journal of the Marine Biological Association of the UK</i> , 81(06), 897-901.	Cooper, W. S. ve Maron, M. E. (1978). Foundations of probabilistic and utility-theoretic indexing. <i>Journal of the ACM (JACM)</i> , 25(1), 67-80.
2	Lucia, L., Lo, D., Jiang, L., Thung, F. ve Budi, A. (2014). Extended comprehensive study of association measures for fault localization. <i>Journal of Software: Evolution and Process</i> , 26(2), 172-219.	Robertson, S. E., Maron, M. E., & Cooper, W. S. (1982). Probability of relevance: A unification of two competing models for document retrieval. <i>Information Technology: Research and Development</i> , 1(1), 1-21.
3	Todeschini, R., Consonni, V., Xiang, H., Holliday, J., Buscema, M. ve Willett, P. (2012). Similarity coefficients for binary chemoinformatics data: Overview and extended comparison using simulated and real data sets. <i>Journal of Chemical Information and Modeling</i> , 52(11), 2884-2901.	Robertson, S. E. (1997). The probability ranking principle in IR. <i>Readings in information retrieval</i> , 281-286.
4	Ghosh, S., Carranza, E. J. M., Van Westen, C. J., Jetten, V. G. ve Bhattacharya, D. N. (2011). Selecting and weighting spatial predictors for empirical modeling of landslide susceptibility in the Darjeeling Himalayas (India). <i>Geomorphology</i> , 131(1), 35-56.	Robertson, S. E. ve Sparck Jones, K. (1976). Relevance weighting of search terms. <i>Journal of the American Society for Information Science</i> , 27(3), 129-146.
5	Tan, P. N., Kumar, V. ve Srivastava, J. (2004). Selecting the right objective measure for association analysis. <i>Information Systems</i> , 29(4), 293-313.	Croft, W. B., & Harper, D. J. (1979). Using probabilistic models of document retrieval without relevance information. <i>Journal of Documentation</i> , 35(4), 285-295.
6	Shoukri, M. M., Chaudhary, M. A. ve Mohamed, G. H. (2003). Evaluating normal approximation confidence intervals for measures of 2x 2 association with applications to twin data. <i>Biometrical Journal</i> , 45(1), 20-33.	Robertson, S. E. (1994). Documentation note query-document symmetry and dual models. <i>Journal of Documentation</i> , 50(3), 233-238. Robertson, S. E. (1994). Documentation note query-document symmetry and dual models. <i>Journal of Documentation</i> , 50(3), 233-238.
7	Reboussin, B. A. ve Liang, K. Y. (1998). An estimating equations approach for the LISCOMP model. <i>Psychometrika</i> , 63(2), 165-182.	Salton, G., Fox, E. A. ve Wu, H. (1983). Extended Boolean information retrieval. <i>Communications of the ACM</i> , 26(11), 1022-1036.
8	Izmaylov, D. M. ve Obukhova, L. K. (1996). Geroprotector efficiency depends on viability of control population: Life span investigation in D - melanogaster. <i>Mechanisms of Ageing and Development</i> , 91(3), 155-164.	Bookstein, A. ve Kraft, D. (1977). Operations research applied to document indexing and retrieval decisions. <i>Journal of the ACM</i> , 24(3), 418-427.
9	Goodman, L. A. (1981). Association models and the bivariate normal for contingency tables with ordered categories. <i>Biometrika</i> , 68(2), 347-355.	Winkler, R. L. (1968). The consensus of subjective probability distributions. <i>Management Science</i> , 15(2), B-61- B-75.
10	Wilson E.B. (1964). Comparative Experiment + Observed Association .2. <i>Proceedings of The National Academy of Sciences of The United States of America</i> . 51(4), 539-542.	Cooper, W. S. (1978). Indexing documents by gedanken experimentation. <i>Journal of the American Society for Information Science</i> , 29(3), 107-119.
11	Evans, F. B. (1959). Psychological and objective factors in the prediction of brand choice ford versus chevrolet. <i>The Journal of Business</i> , 32(4), 340-369.	Bookstein, A. (1983). Outline of a general probabilistic retrieval model. <i>Journal of Documentation</i> , 39(2), 63-72.
12	Li, F., Chu, H., ve Nie, L. (2011). A two-stage estimation for screening studies using two diagnostic tests with binary disease status verified in test positives only. <i>Statistical Methods in Medical Research</i> , 24(6), 635-56.	Bookstein, A. ve Cooper, W. (1976). A general mathematical model for information retrieval systems. <i>The Library Quarterly</i> , 153-167.
13	Crusto, C. A., Dantzier, J., Roberts, Y. H., & Hooper, L. M. (2015). Psychometric Evaluation of Data From the Race-Related Events Scale. <i>Measurement and Evaluation in Counseling and Development</i> , 48(4), 285-296.	Luhn, H. P. (1961). The automatic derivation of information retrieval encodings from machine-readable texts. A. Kent (Yay. haz). <i>Information retrieval and machine translation</i> (Cilt 3, Bölüm 2, ss. 1021-1028) içinde. New York: Interscience Publication.
14	Alofs, K. M. ve Jackson, D. A. (2015). The vulnerability of species to range expansions by predators can be predicted using historical species associations and body size. <i>Proceedings of the Royal Society of London B: Biological Sciences</i> , 282(1812). doi: 10.1098/rspb.2015.1211	Fuhr, N. (1989). Models for retrieval with probabilistic indexing. <i>Information Processing and Management</i> , 22(1), 55-72.
15	Baker, S. G. (2015). RE: Combined Associations of Genetic and Environmental Risk Factors: Implications for Prevention of Breast Cancer. <i>Journal of the National Cancer Institute</i> , 107(6), djv127.	Bookstein, A. (1978). On the perils of merging Boolean and weighted retrieval systems. <i>Journal of the American Society for Information Science</i> , 29(3), 156-158.
16	Yousefi, M. ve Carranza, E. J. M. (2015). Prediction-area (P-A) plot and C-A fractal analysis to classify and evaluate evidential maps for mineral prospectivity modeling. <i>Computers and Geosciences</i> , 79, 69-81.	Croft, W. B. ve Lafferty, J. (Yay. haz.). (2003). <i>Language modeling for information retrieval</i> . Boston: Kluwer Academic.
17	Negri, M. ve Sprumont, Y. (2015). Size invariant measures of association: Characterization and difficulties. <i>Mathematical Social Sciences</i> , 75, 115-122.	Bookstein, A. (1979). Relevance. <i>Journal of the American Society for Information Science</i> , 30(5), 269-273.



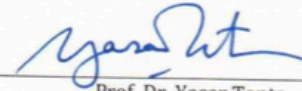
Sıra No	WoS Kaynakça Benzerliğine Göre Oluşturulan İlgililik Sıralaması	Pennant Diyagramları Yöntemine Göre Oluşturulan İlgililik Sıralaması
18	Lordan, J., Alegre, S., Moerkens, R., Sarasúa, M. J. ve Alins, G. (2015). Phenology and interspecific association of <i>Forficula auricularia</i> and <i>Forficula pubescens</i> in apple orchards. <i>Spanish Journal of Agricultural Research</i> , 13(1), 1003.	Swanson, D. R. (1960). Searching natural language text by computer. <i>Science</i> , 132(3434), 1099-1104.
19	Otte, S., Lang, F. U., Rasche, K., Bernheim, D., Vasic, N. ve Dudeck, M. (2015). Convergent validity of SCID-II and TCI: categorical vs. dimensional diagnostics of personality disorders using the example of short-term prisoners. <i>Fortschritte der Neurologie-Psychiatrie</i> , 83(3), 157-161.	Maron, M. E. (1977). On indexing, retrieval and the meaning of about. <i>Journal of the American Society for Information Science</i> , 28(1), 38–43.
20	Allen, M. L., Wallace, C. F. ve Wilmers, C. C. (2015). Patterns in bobcat ( <i>Lynx rufus</i> ) scent marking and communication behaviors. <i>Journal of ethology</i> , 33(1), 9-14.	Turtle, H. R ve Croft, W. B. (1991). Evaluation of an inference network-based retrieval model. <i>ACM Transactions on Information Systems</i> , 9(3), 187–222.
21	Dusek, L., Pavlik, T., Jarkovsky, J. ve Koptikova, J. (2015). Analysis of data in neurology LI. Yules Q-a usefull tool for comparing the various estimates of the ratio of chances. <i>Casks a Slovenska Neurologie a Neurochirurgie</i> , 78(3), 367-369.	Miller (1971). A probabilistic search strategy for Medlars. <i>Journal of Documentation</i> , 27, 254–266.
22	Egozcue, J. J., Pawlowsky-Glahn, V., Templ, M. ve Hron, K. (2015). Independence in contingency tables using simplicial geometry. <i>Communications in Statistics-Theory and Methods</i> , 44(18), 3978-3996.	Thompson, P. (1988). Subjective probability and information retrieval: A review of the psychological literature. <i>Journal of Documentation</i> , 44(2), 119–143.
23	Egozcue, J. J., Pawlowsky-Glahn, V., Templ, M. & Hron, K. (2015). Independence in contingency tables using simplicial geometry. <i>Communications in Statistics-Theory and Methods</i> , 44(18), 3978-3996.	Bodoff, D. ve Robertson, S. E. (2004). New unified probabilistic model. <i>Journal of the American Society for Information Science and Technology</i> , 55(6), 471–487.
24	Xu, S. ve Lorber, M. F. (2014). Interrater agreement statistics with skewed data: Evaluation of alternatives to Cohen's kappa. <i>Journal of consulting and clinical psychology</i> , 82(6), 1219.	Salton, G., Yang, C. S. ve Yu, C. T. (1975). A theory of term importance in automatic text analysis. <i>Journal of the American Society for Information Science</i> , 26(1), 33-44.
25	Liao, S. G., Lin, Y., Kang, D. D., Chandra, D., Bon, J., Kaminski, N. ve Tseng, G. C. (2014). Missing value imputation in high-dimensional phenomic data: Imputable or not, and how?. <i>BMC bioinformatics</i> , 15(1), 1.	Bookstein, A. (1985). Probability and fuzzy-set applications to information retrieval. <i>Annual Review of Information Science and Technology</i> , 20, 117-151.
26	Keiding, N. ve Clayton, D. (2014). Standardization and control for confounding in observational studies: A historical perspective. <i>Statistical Science</i> , 29(4), 529-558.	Saracevic, T. (1975). Relevance: A review of and a framework for thinking on the notion in information science. <i>Journal of the American Society for Information Science</i> , 26, 321–343.
27	Cottler, L. B., Ajinkya, S., Goldberger, B. A., Ghani, M. A., Martin, D. M., Hu, H. ve Gold, M. S. (2014). Prevalence of drug and alcohol use in urban Afghanistan: Epidemiological data from the Afghanistan National Urban Drug Use Study (ANUDUS). <i>The Lancet Global Health</i> , 2(10), e592-e600.	Wong, S. K. M. ve Yao, Y. Y. (1995). On modeling information retrieval with probabilistic inference. <i>ACM Transactions on Information Systems (TOIS)</i> , 13(1), 38-68.
28	Teegavarapu, R. S. (2014). Missing precipitation data estimation using optimal proximity metric-based imputation, nearest-neighbour classification and cluster-based interpolation methods. <i>Hydrological Sciences Journal</i> , 59(11), 2009-2026.	Maron, M. E., Curry, S. ve Thompson, P. (1986). An inductive search system: Theory, design and implementation. <i>IEEE Transactions on Systems, Man, and Cybernetics</i> , SMC-16(1), 21–28
29	Balasubramanian, R., Andres Houseman, E., Coull, B. A., Lev, M. H., Schwamm, L. H. ve Betensky, R. A. (2014). Variable importance in matched case-control studies in settings of high dimensional data. <i>Journal of the Royal Statistical Society: Series C (Applied Statistics)</i> , 63(4), 639-655.	Bush, V. (1945). As we may think. <i>Atlantic Monthly</i> , 176(1), 101-108.
30	Tew, C., Giraud-Carrier, C., Tanner, K. ve Burton, S. (2014). Behavior-based clustering and analysis of interestingness measures for association rule mining. <i>Data Mining and Knowledge Discovery</i> , 28(4), 1004-1045.	Blair, D. C. ve Maron, M. E. (1985). An evaluation of retrieval effectiveness for a full-text document-retrieval system. <i>Communications of the ACM</i> , 28(3), 289–299.
31	Ramkumar, T., Srinivasan, R. ve Hariharan, S. (2014). Synthesizing Global Association Rules from Different Data Sources Based on Desired Interestingness Metrics. <i>International Journal of Information Technology and Decision Making</i> , 13(03), 473-495.	Salton, G. ve Yang, C. S. (1973). On the specification of term values in automatic indexing. <i>Journal of Documentation</i> , 29, 351–372.
32	Cicchetti, D. V. (2014). On scales of measurement in autism spectrum disorders (ASD) and beyond: Where Smitty went wrong. <i>Journal of Autism and Developmental Disorders</i> , 44(2), 303-309.	Dos Santos, E. B., Hruschka E. R., Hruschka, E. R. ve Ebecken, N. F. (2011). Bayesian network classifiers: Beyond classification accuracy. <i>Intelligent Data Analysis</i> , 15(3), 279-298.
33	Allen, M. L., Wittmer, H. U. ve Wilmers, C. C. (2014). Puma communication behaviours: understanding functional use and variation among sex and age classes. <i>Behaviour</i> , 151(6), 819-840.	Gordon, M. (1988). Probabilistic and genetic algorithms in document retrieval. <i>Communications of the ACM</i> , 31(10), 1208-1218.
34	Beh, E. J., Tran, D. ve Hudson, I. L. (2013). A reformulation of the aggregate association index using the odds ratio. <i>Computational Statistics and Data Analysis</i> , 68, 52-65.	Swanson, D. R. (1977). Information retrieval as a trial-and-error process. <i>The Library Quarterly</i> , 128-148.

Sıra No	WoS Kaynakça Benzerliğine Göre Oluşturulan İlgilik Sıralaması	Pennant Diyagramları Yöntemine Göre Oluşturulan İlgilik Sıralaması
35	Cui, Y. ve Honkala, M. (2013). A novel mobile device user interface with integrated social networking services. <i>International Journal of Human-Computer Studies</i> , 71(9), 919-932.	Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. <i>Journal of Documentation</i> 28(1): 11–21. doi: 10.1108/eb026526.
36	Malik, S. ve Singh, R. (2013). An improved estimator using two auxiliary attributes. <i>Applied mathematics and computation</i> , 219(23), 10983-10986.	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, Part 1: The conceptual model. <i>Information Processing and Management</i> , 26(3), 371–382.
37	Shen, S. Z., Zhang, H., Shi, G. R., Li, W. Z., Xie, J. F., Mu, L. ve Fan, J. X. (2013). Early Permian (Cisuralian) global brachiopod palaeobiogeography. <i>Gondwana Research</i> , 24(1), 104-124.	Fuhr, N. (1986). Two models of retrieval with probabilistic indexing. F. Rabitti (Yay. haz.), <i>ACM Conference on research and development in information retrieval (s. 249–257)</i> , 8–10 September 1986, Pisa, Italy.
38	Morelli, J. N.ve Bokhari, D. (2013). Distribution of scholarly publications among academic radiology departments. <i>Journal of the American College of Radiology</i> , 10(3), 190-194.	Van Rijsbergen, C. J., Harper, D. J. ve Porter, M. F. (1981). The selection of good search terms. <i>Information Processing and Management</i> , 17(2), 77-91.
39	Hammond, S. (2013). <i>Introducing a Common Range Index of Inter-variable Similarity for the Analysis of Radex Structures</i> .	Bodoff, D. (1999). A re-unification of two competing models for document retrieval. <i>Journal of the American Society for Information Science</i> , 50(1), 49–64.
40	Heynderickx, P. M., Van Huffel, K., Dewulf, J. ve Van Langenhove, H. (2013). Application of similarity coefficients to SIFT-MS data for livestock emission characterization. <i>Biosystems engineering</i> , 114(1), 44-54.	Thompson, P. (1990). A combination of expert opinion approach to probabilistic information retrieval, Part 2: Mathematical treatment of CEO model 3. <i>Information Processing and Management</i> , 26(3), 383–394.
41	Unnewehr, S., Schneider, S. ve Margraf, J. (1995). <i>Diagnostisches Interview bei psychischen Störungen im Kindes-und Jugendalter (Kinder-DIPS)</i> . Heidelberg: Springer.	Cleverdon, C.W. (1962). <i>Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems</i> . Cranfield College of Aeronautics,.
42	Miklós-Thal, J. (2012). Linking reputations through umbrella branding. <i>Quantitative Marketing and Economics</i> , 10(3), 335-374.	Salton, G., Wong, A. ve Yang, C. S. (1975). A vector space model for automatic indexing. <i>Communications of the ACM</i> , 18 (11), 613-620.
43	Albatineh, A. N., Khan, H. M. ve Niewiadomska-Bugaj, M. (2012). On the equivalence of some indices of similarity: Implication for binary presence/absence data. <i>Australian ve New Zealand Journal of Statistics</i> , 54(2), 189-198.	Salton, G. ve Lesk, M. E. (1968). Computer evaluation of indexing and text processing. <i>Journal of the ACM</i> , 15, 8–36.
44	Aguinis, H. (2012). The best and the rest: Revisiting the norm of normality of individual performance. <i>Personnel Psychology</i> , 65(1), 79-119.	Luk, R. W. (2008). On event space and rank equivalence between probabilistic retrieval models. <i>Information Retrieval</i> , 11(6), 539-561.
45	Duffy, K. R., Wellard, C. J., Markham, J. F., Zhou, J. H., Holmberg, R., Hawkins, E. D. ve Hodgkin, P. D. (2012). Activation-induced B cell fates are selected by intracellular stochastic competition. <i>Science</i> , 335(6066), 338-341.	Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. <i>Proceedings of the ACM-SIAM Symposium on discrete algorithms</i> (s.1-33) içinde.
46	Baraniuk, S., Seay, R., Sinha, A. K. ve Piller, L. B. (2012). Comparison of the Global Statistical Test and Composite Outcome for Secondary Analyses of Multiple Coronary Heart Disease Outcomes. <i>Progress in Cardiovascular Diseases</i> , 54(4), 357-361.	Maron, M. E. (1984). Probabilistic retrieval models. B. Dervin ve M. J. Voigt (Yay. haz.), <i>Progress in communication sciences</i> (s. 145–176) içinde.
47	Heynderickx, P. M., Van Huffel, K., Dewulf, J. ve Van, H. (2012). SIFT-MS for livestock emission characterization: application of similarity coefficients. <i>Chemical Engineering</i> , 30, 157-162.	Porter, M. F. (1980). An algorithm for suffix stripping. <i>Program (Automated Library and Information Systems)</i> , 14(3), 130-137.
48	Mair, P., Satorra, A. ve Bentler, P. M. (2012). Generating nonnormal multivariate data using copulas: Applications to SEM. <i>Multivariate Behavioral Research</i> , 47(4), 547-565.	Salton, G., Wong, A. ve Yu, C. T. (1976). Automatic indexing using term discrimination and term precision measurements. <i>Information Processing and Management</i> , 12(1), 43-51.
49	Saur, R., Maier, C., Milian, M., Riedel, E., Berg, D., Liepelt-Scarfone, I. ve Leyhe, T. (2012). Clock test deficits related to the global cognitive state in Alzheimer's and Parkinson's disease. <i>Dementia and Geriatric Cognitive Disorders</i> , 33(1), 59-72.	Chow, C. ve Liu, C. (1968). Approximating discrete probability distributions with dependence trees. <i>IEEE Transactions on Information Theory</i> , 14(3), 462-467.
50	Teegavarapu, R. S. (2012). <i>Floods in a changing climate: extreme precipitation</i> . Cambridge University Press.	Croft, W. B. (2000). Combining approaches to information retrieval. W. B. Croft (Yay. haz.), <i>Advances in information retrieval: Recent research from the center for intelligent information retrieval</i> (s.1-36) içinde. Boston: Kluwer.

## Ek 4: Orijinallik Raporu

 <p><b>HACETTEPE ÜNİVERSİTESİ</b> <b>SOSYAL BİLİMLER ENSTİTÜSÜ</b> <b>YÜKSEK LİSANS/DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU</b></p>
<p><b>HACETTEPE ÜNİVERSİTESİ</b> <b>SOSYAL BİLİMLER ENSTİTÜSÜ</b> <b>BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞI'NA</b></p> <p style="text-align: right;">Tarih: 23/06/2016</p> <p>Tez Başlığı : Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi</p> <p>Yukarıda başlığı gösterilen tez çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 87 sayfalık kısmına ilişkin, 20/06/2016 tarihinde tez danışmanım tarafından Turnitin adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 1'dir.</p> <p>Uygulanan filtrelemeler:</p> <ol style="list-style-type: none"> <li>1- Kabul/Onay ve Bildirim sayfaları hariç,</li> <li>2- Kaynakça hariç</li> <li>3- Alıntılar hariç</li> <li>4- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç</li> </ol> <p>Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.</p> <p>Gereğini saygılarımla arz ederim.</p> <p style="text-align: right;">23.06.2016  Tarih ve İmza</p> <p><b>Adı Soyadı:</b> Müge Akbulut <b>Öğrenci No:</b> N11123286 <b>Anabilim Dalı:</b> Bilgi ve Belge Yönetimi <b>Statüsü:</b> <input checked="" type="checkbox"/> Y.Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.</p>
<p><b><u>DANIŞMAN ONAYI</u></b></p> <p style="text-align: center;">UYGUNDUR.</p> <p style="text-align: center;"> Prof. Dr. Yaşar Tonta</p>

## Ek 5: Etik Kurul Muafiyet İzni

 <p><b>HACETTEPE ÜNİVERSİTESİ</b> <b>SOSYAL BİLİMLER ENSTİTÜSÜ</b> <b>TEZ ÇALIŞMASI ETİK KURUL İZİN MUAFİYETİ FORMU</b></p>
<p><b>HACETTEPE ÜNİVERSİTESİ</b> <b>SOSYAL BİLİMLER ENSTİTÜSÜ</b> <b>BİLGİ VE BELGE YÖNETİMİ ANABİLİM DALI BAŞKANLIĞI'NA</b></p> <p style="text-align: right;">Tarih: 23/06/2016</p> <p>Tez Başlığı: Atıf Klasiklerinin Etkisinin ve İlgililik Sıralamalarının Pennant Diyagramları ile Analizi</p> <p>Yukarıda başlığı gösterilen tez çalışmam:</p> <ol style="list-style-type: none"> <li>1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,</li> <li>2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.</li> <li>3. Beden bütünlüğüne müdahale içermemektedir.</li> <li>4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.</li> </ol> <p>Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.</p> <p>Gereğini saygılarımla arz ederim.</p> <p style="text-align: right;">23.06.2016  Tarih ve İmza</p> <p><b>Adı Soyadı:</b> Müge Akbulut <b>Öğrenci No:</b> N11123286 <b>Anabilim Dalı:</b> Bilgi ve Belge Yönetimi <b>Programı:</b> Bilgi ve Belge Yönetimi <b>Statüsü:</b> <input checked="" type="checkbox"/> Y.Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.</p>
<p><b><u>DANIŞMAN GÖRÜŞÜ VE ONAYI</u></b></p> <p style="text-align: center;"> Prof. Dr. Yaşar Tonta</p> <p><b>Detaylı Bilgi:</b> <a href="http://www.sosyalbilimler.hacettepe.edu.tr">http://www.sosyalbilimler.hacettepe.edu.tr</a> <b>Telefon:</b> 0-312-2976860 <b>Faks:</b> 0-3122992147 <b>E-posta:</b> <a href="mailto:sosyalbilimler@hacettepe.edu.tr">sosyalbilimler@hacettepe.edu.tr</a></p>